11-17-2017

Achilles Is a Circadian Clock-controlled Gene That Regulates the Immune System and Its Rhythmicity in Drosophila

Jiajia Li
jl9q6@umsl.edu

ACHILLES IS A CIRCADIAN CLOCK-CONTROLLED GENE THAT REGULATES THE IMMUNE SYSTEM AND ITS RHYTHMICITY IN *DROSOPHILA*

by

Jiajia Li

Master of Science –Graduate, Shanghai Institute of Biochemistry and Cell Biology, 2011
Bachelor of Science – Undergraduate, China Agricultural University, 2008

A Dissertation

Submitted to The Graduate School of

University of Missouri-St. Louis

In partial fulfillment of the requirements for the degree

Doctor of Philosophy

In

Biology

With an emphasis in

Cell and Molecular Biology

December 2017

Advisory Committee

Sam Wang, Ph.D.
Chairperson

Lon Chubiz, Ph.D.

Aimee Dunlap, Ph.D.

Michael E. Hughes, Ph.D.

Table of Contents

Abstract

Circadian clock is a transcriptional/translational feedback loop that drives the rhythmic expression of downstream mRNAs. Termed "clock-controlled genes," these molecular outputs of the circadian clock orchestrate cellular, metabolic, and behavioral rhythms. As part of our on-going work to characterize key upstream regulators of circadian mRNA expression, we have identified a novel clock-controlled gene in *Drosophila melanogaster*, *Achilles* (*Achl*), which is rhythmic at the mRNA level in the brain and represses expression of immune response genes, especially anti-microbial peptides in the immune system. *Achl* knock-down in the brain dramatically elevates expression of crucial immune response genes, including *IM1* (*Immune induced molecule 1*), *Mtk* (*Metchnikowin*), and *Drs* (*Drosomysin*). As a result, flies with knocked-down *Achl* expression are more resistant to bacterial challenges. Meanwhile, no significant change in core clock gene expression and locomotor activity is observed, suggesting that *Achl* influences rhythmic mRNA outputs rather than directly regulating the core timekeeping mechanism. Additionally, *Achl* knock-down in the brain disrupts the rhythmicity of the immune system. Flies with knocked-down *Achl* show altered rhythmicity in both survival towards infection and sensitivity of immune response gene induction upon infection. Using high-throughput RNA-sequencing, we also identified candidate clock controlled genes that are downstream of *Achl*. Notably, *Achl* knock-down in the absence of immune challenge significantly diminishes the fly's overall lifespan and resistance towards starvation, indicating a behavioral or metabolic cost of constitutively activating this pathway. Together, our data demonstrate that (1) *Achl* is a novel clock-controlled gene that (2) regulates the immune system in a repressive manner. (3) *Achl* regulates the rhythmicity of the immune system, and (4) *Achl* participates in signaling from the brain to immunological tissues.

Acknowledgements

First of all, I would like to thank my advisor, Dr. Michael Hughes for his support and trust throughout my Ph.D. training process. I really appreciate his hands-off yet approachable mentoring style. His critical thinking and enthusiasm towards research is always inspiring. In addition to research, he also helped a lot with my writing and presenting skills and career development. The training I got from Dr. Hughes will have a life-long impact for me.

I would also thank my dissertation committee members for their help and suggestions throughout my graduate study. Besides, I would like to thank my comprehensive exam committee members for their help and suggestions.

I would also thank the entire Hughes Lab, especially Erin Terry, for their support, help, and discussions during the past years. I want to thank our Biology Department as well. It is a great community that we learn from each other and have fun together. I appreciate the training opportunities the department provides for graduate students. I would thank the colleagues and classmates for their inspiring discussion and suggestions. I would especially thank Dr. Thiel lab and Dr. Chubiz lab for sharing equipment and reagents.

I would like to thank the graduate school, for providing training resources and funds. I am grateful to the dissertation fellowship that supports the completion of my dissertation.

I would also thank colleagues in the pulmonary and critical care division at Washington University in St. Louis during my last year of research. Special thanks to Dr. Michael Holtzman lab and Dr. Arthur Romero for sharing equipment.

I would like to thank Dr. Paul Taurget and Xitong Liang, and Dr. James Skeath and Beth Wilson for sharing fly lines. I would like to thank Dr. Tracey Hermanstyne and Dr. Dr. Jeanne Nerbonne for sharing equipment.

I would like to thank my collaborators for their smart ideas and hard work. The project won't come to a good story without their efforts.

I would also like to thank my previous mentors, especially Dr. Degui Chen, for their training and support.

I also want to thank my parents for their love and support throughout my life. They are always there to encourage me every time I get frustrated. And it is them who give me the strength to realize my potential and chase my dreams. I would also thank other family members for their love and support.

Lastly, I would like to thank my husband, Dr. Chao Wu. His support and encouragement helped me go through the hard days. It is his love and accompany that give me the strength to pursue my Ph.D. in U.S, half a planet away from my home country.

Dedicated to my parents.

List of Figures

Chapter 1

Chapter 2

Chapter 3

List of Tables

Chapter 3

Chapter 4

Chapter 1: Introduction

Outline

There will be five chapters in this Dissertation. In Chapter 1, I will introduce the basic background information and related research progress. In Chapter 2, I will focus on a computational simulation and further analysis that was done to shed light on the considerations when using RNA-sequencing technology to profile rhythmic mRNA expression. Results obtained from this simulation also contribute to the experiments discussed in the Chapter 4. In Chapters 3 & 4, I will talk about the major dissertation project about the role of *AchI* gene in the regulation of the immune function and its circadian rhythms in *Drosophila* (fruit flies). In Chapter 3, I will focus on the role *AchI* plays in the regulation of the immune system and in Chapter 4, I will focus on the circadian aspect of the immune system and the role *AchI* plays in its regulation. There will be a more specific introduction about the background of the project described in Chapter 3 and 4 in Chapter 3. In the last chapter, Chapter 5, I will make an overall conclusion and future directions.

Circadian Rhythms

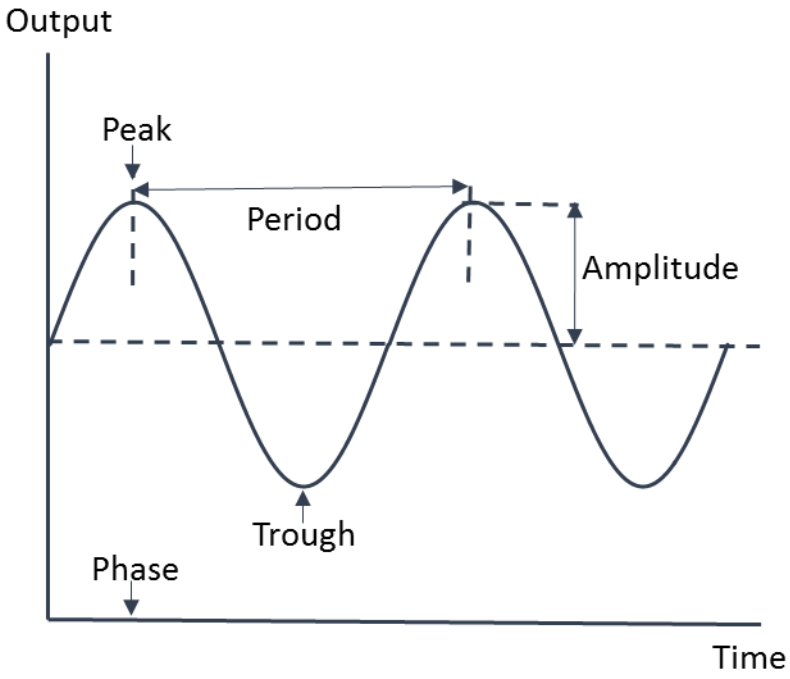
Our planet completes one rotation roughly every 24 hours, giving us day and night. As a consequence, most organisms, including bacteria, fungi, plants and animals, have developed an endogenous clock system to predict and adapt to the predictable light and dark cycle. This internal clock system is named circadian rhythms, with “circa” referring to “approximately”, and “dies” referring to “day” in *Latin*. Circadian rhythms participate in almost all aspects of physiology and maintain their homeostasis. In mammals, for example, locomotor activity, sleep-wake cycle, feeding, hormone releasing, body temperature, blood pressure and cardiovascular activity, muscle strength, metabolism and alertness are all rhythmic. Disruption of circadian rhythms has been implicated in multiple pathologies, like neurodegenerative disease, cardiovascular disease,

obesity, diabetes, metabolic syndromes, cancer and depression (Klerman 2005, Halberg et al. 2006, Levi and Schibler 2007).

There are three key characteristics for a circadian rhythm: period, phase and amplitude. As illustrated in Figure 1.1, period is the time elapsed for one complete cycle, phase is the time when a rhythm reaches its peak, and amplitude is the difference between peak (or trough) to the mean level. Period and phase defines the property of specific rhythm, while amplitude defines its rhythmic strength.

At the organismal level, here are three features within the circadian rhythms. First, circadian rhythms can be entrained. When there are environmental cues, named “zeitgebers”, animals are able to synchronize their rhythms with various zeitgebers, among which light is the strongest zeightber. Second, circadian rhythms are self-sustained. Under constant conditions, where no environmental cue is available, animals are still able to maintain their rhythmicity, though the amplitude is damped. We use different terms to identify time so that we can easily distinguish the environmental conditions of experimental objects. Under constant conditions, a standard of time, termed circadian time (CT) is used. CT0 is defined as the onset of activity of diurnal organisms, that is, subjective dawn. CT12 is defined as the onset of activity of nocturnal organisms, that is, subjective dusk. When an environmental cue, like light that is most commonly used in lab, is available, zeitgeber time (ZT) is used. ZT0 is defined as the onset of light, and ZT12 is defined as the offset of light. Third, the circadian rhythms are temperature compensated. That is, animals are able to maintain their period length over any constant temperature within the physiological temperature range. For example, fruit flies of the same genetic background maintained at either 18 degrees or 25 degrees will have the same period length.

Figure 1.1. Illustration of the key characteristics of a rhythm



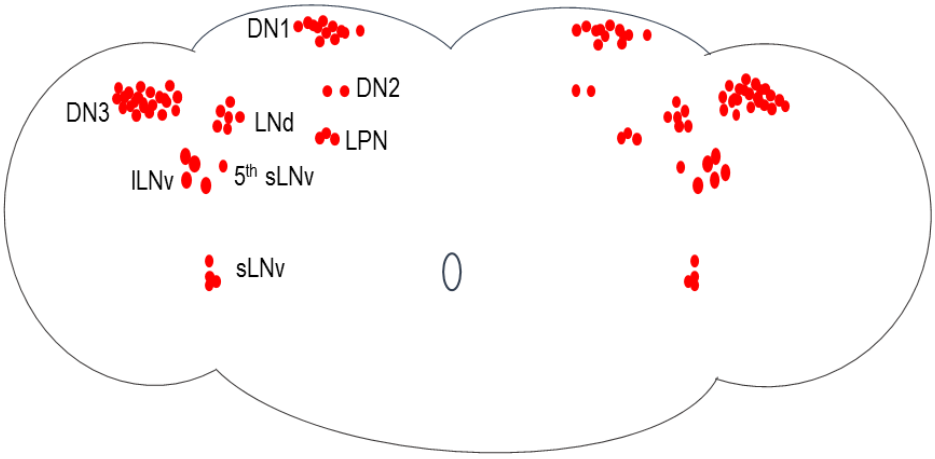
For a defined circadian rhythm output, there is a peak where the output reaches its highest level, and a trough where the output reaches its lowest level. Period is the time elapsed for an entire cycle, which will be around 24 hours for circadian rhythms. Phase is the time when the output reaches its peak. And the amplitude is the range of output between peak (or trough) to the median.

Circadian rhythms are highly hierarchical

Regulation of the circadian rhythms is highly hierarchical. The principal oscillator (also called master pacemaker) in mammals is the suprachiasmatic nucleus (SCN) located at the anterior hypothalamus in the brain (Hastings et al. 2003). There are about 20,000 cells within the SCN. These SCN cells receive both optic and non-optic signals from the environment. The most potent input signal to the SCN is light, which is received by rods, cones and intrinsically photosensitive retinal ganglion cells (iPRGCs) within the eyes (Schmidt et al. 2011). The received light signal is transmitted to the SCN through the photic neural input pathway, retino-hypothalamic tract (RHT) (Slat et al. 2013). SCN cells incorporate these input signals received from the environment, synchronize the signals within SCN cells, and transmit synchronized output signals to other brain regions and peripheral tissues. Output signals are transmitted through neuronal connections and secretary humoral factors, like TGF-α, prokineticin and cortisol. SCN cells ultimately coordinate rhythmic behavior, metabolism and physiology in peripheral tissues through this signal transmitting system (Slat et al. 2013).

Though peripheral tissues mainly synchronize through SCN released signals, rhythms in peripheral tissues are endogenous and self-sustaining. Peripheral tissues possess the same functional molecular components that are involved in SCN rhythms and can keep rhythmicity in vitro, though not as robust as SCN rhythms (Yoo et al. 2004). Furthermore, isolated cells can maintain rhythmicity for years in vitro, suggesting that peripheral tissues are able to sustain circadian rhythms autonomously (Balsalobre et al. 2000, Nagoshi et al. 2004). Peripheral oscillators can also receive signals from the local environment and entrain their rhythms accordingly. For example, restricted-feeding can cause a rhythmic shift in the liver, without affecting the SCN rhythm (Kornmann et al. 2007).

Figure 1.2. The clock neuron network in Drosophila



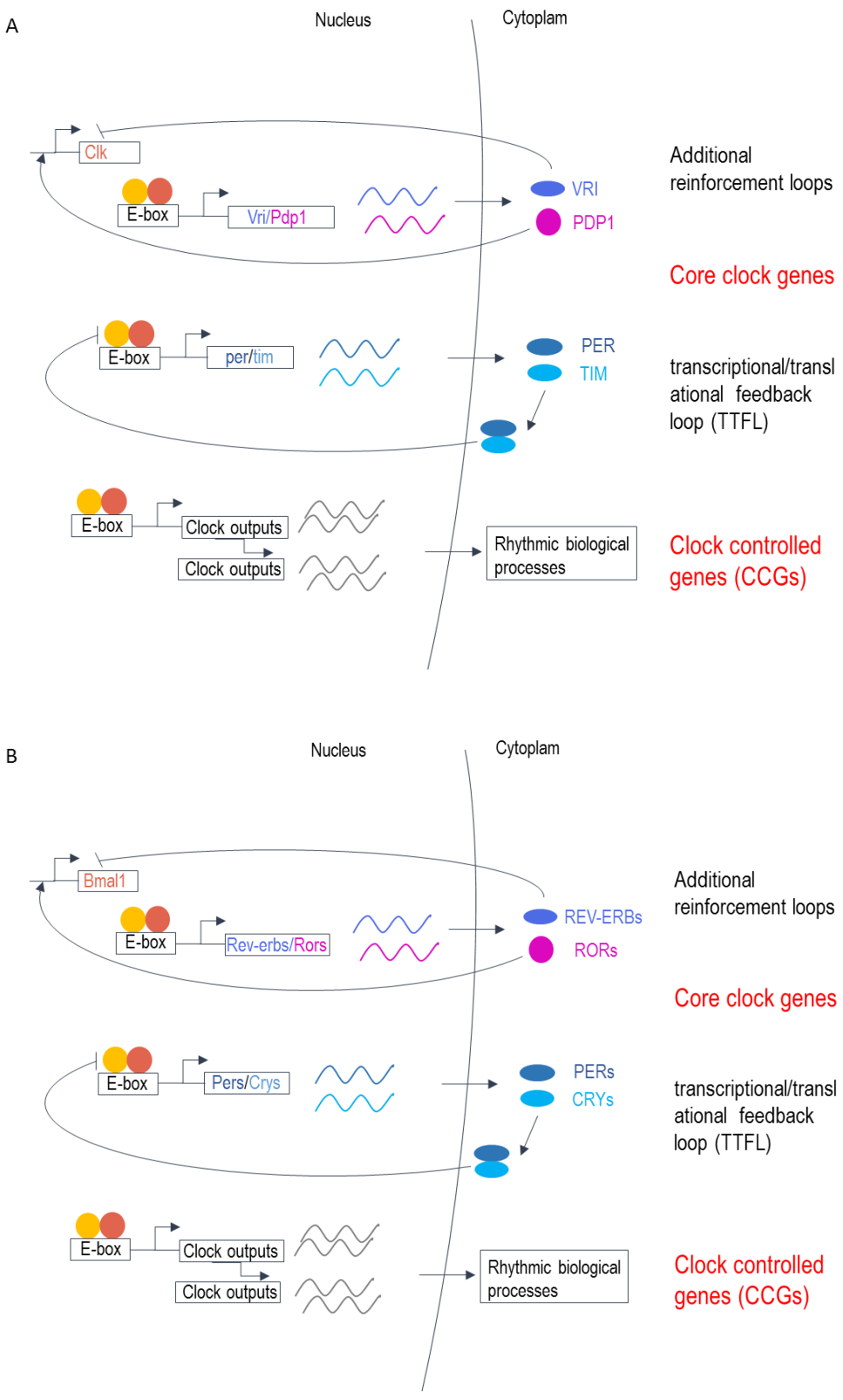
Schematic figure showing the location of principal oscillator neurons in flies. There are about 150 clock neurons that are divided into several group based on their clustering pattern, location and size. This figure is modified from Allada & Chung. (2010) Annual Review of Physiology.

In *Drosophila*, the principal oscillator network is composed of about 150 neurons (out of about 250,000 neurons) that express the core clock genes (Figure 1.2). These 150 neurons, including dorsal lateral neurons (LNd), small and large ventral lateral neurons (lLNv and sLNv), lateral-posterior neurons (LPN) and dorsal neurons (DN1, DN2 and DN3), are the functional compartment of SCN neurons in *Drosophila* (Nitabach and Taghert 2008, Allada and Chung 2010). These cells receive light input signals from the retina, ocelli, and the Hofbauer-Buchner eyelets. In addition, CRYPTOCHROME (CRY), a blue-light photoreceptor, provides an alternative light input pathway as CRY is activated upon light, and the activation of CRY will promote TIM degradation and thus synchronize the clock with the environmental light.

Molecular basis of the circadian clock

At the molecular level, circadian rhythms are regulated by core clock genes that compose self-sustained 24-hour feedback loops (Ko and Takahashi 2006). In flies, as shown in Figure 1.3, the structure of the molecular clock is conserved between flies (Figure 1.3A) and mice (Figure 1.3B). Using fly molecular clock for example, two transcription factors, CLOCK (CLK) and CYCLE (CYC) form heterodimers and bind to the E-box element (CANNTG), located at promoter regions of *period* (*per*) and *timeless* (*tim*), promoting their expression. PER and TIM, once expressed, will dimerize and translocate into the nucleus, where the heterodimers prevent CLK and CYC heterodimers from accessing E-box elements, thus decreasing the expression of per and tim themselves. When the levels of PER and TIM get low, they will be degraded in the cytoplasm before translocating into the nucleus. The inhibition of CLK and CYC activity will thus be released. When CLK and CYC activities get released, expression of *per* and *tim* will be activated again (Allada and Chung 2010). This transcriptional-translational feedback loop (TTFL) occurs

Figure 1.3. Molecular components of the clock in flies and mice



The molecular architecture of the clock system, including the core clock genes and CCGs, is conserved in flies (A) and mice (B). This figure is modified from Ko & Takahashi. (2006) Human Molecular Genetics.

every 24 hours, with some delay mechanisms that are controlled by protein modifications, constitutes the fundamental basis of circadian rhythms (Dunlap 1999).

This TTFL is reinforced by additional loops composed of PDP1 (Par Domain Protein 1)/VRI (Vrille) and CWO (Clockwork Orange). As shown in Figure 1.4B, the expression of *Pdp1*, *vri* and *cwo*, are under the control of CLK and CYC. Meanwhile, PDP1 is a basic zipper (bZip) activator; it activates *Clk* expression through the P/V element located at the promoter region of *Clk*. VRI is a bZip transcriptional repressor which represses *Clk* expression by preventing this PDP binding. CWO is a bHLH repressor; it specifically binds to E-box regions and prevents the expression of CLK/CYC target mRNAs. However, CWO also plays a role of activator in maintaining the robustness of the rhythmicity since loss of *cwo* results in an overall decreased amplitude (Allada and Chung 2010).

Clock-controlled-genes (CCGs) and physiology

These core clock genes maintain circadian rhythms through TTFLs. Furthermore, they control the expression of thousands of downstream genes, referred to as clock controlled genes (CCGs), through either direct transcriptional regulation or cascaded regulation mediated by regulating CCGs that are controlled by TTFL-like transcription factors. While core clock genes maintain circadian rhythms, CCGs are the ones that are directly involved in physiology. The disruption of core clock genes causes systematic rhythmic disorders, while the disruption of CCGs is more likely to cause local disorders (Ko and Takahashi 2006).

Regulation of CCGs is primarily at the transcriptional level through TTFLs, which are most extensively studied. However, there are multiple levels of CCG regulation, including regulation at transcriptional level, post-transcriptional level, translational level and post-translational level. For

example, peroxiredoxins in blood cells undergo 24-hour redox oscillation at the post-translational level (O'Neill and Reddy 2011).

RNA-sequencing is a great tool to study CCGs

To identify CCGs and understand their regulation mechanisms, microarray and RNA-sequencing (RNA-seq) are being widely used. These high through-put profiling based studies have greatly accelerated our understandings of CCGs in wild type animals, different tissues, cell types, the regulation mechanisms and the contributions of principle oscillator and peripheral oscillators (Harmer et al. 2000, McDonald and Rosbash 2001, Wichert et al. 2004, Wijnen et al. 2005, Keegan et al. 2007, Covington et al. 2008, Hughes et al. 2012, Hsu and Harmer 2014, Zhang et al. 2014). In addition, because of the single base pair resolution, RNA-seq data enable the detection of novel cycling transcripts, as well as the measurement of multiple RNA processing forms, like alternative splicing and RNA editing (Trapnell et al. 2010, Filichkin and Mockler 2012, McGlincy et al. 2012).

Rhythmic outputs and immune response

As mentioned earlier, almost all aspects of physiology are under circadian control. These rhythmic physiological activities are the ultimate outputs under the regulation of circadian clock. Immunological defense is one of the most dramatic examples of a pathway through which the circadian clock influences organismal health and fitness. In mammals, the immune system is regulated by circadian rhythms. This is seen at both a molecular and cellular level (Silver et al. 2012a, Silver et al. 2012b, Curtis et al. 2015).

At the molecular level, core clock genes are expressed rhythmically in all immune tissues and

cells. Besides, microarray based global circadian gene expression profile found that over a thousand genes are rhythmically expressed in macrophage (Keller et al. 2009). In addition, upon infection, many cytokines and chemokines, including IL6 (interleukin 6), TNFα (tumor necrosis factors alpha) and CXCL 12 (Chemokine (C-X-C Motif) Ligand 12) are released into the circulation in a rhythmic manner. At the cellular level, the precursor haematopoietic stem cells enter the circulation in a rhythmic manner. Differentiated lymphatic cells, including T lymphocytes, B lymphocytes, natural killer cells, macrophages, monocytes, also show rhythmic trafficking in the circulation (Mendez-Ferrer et al. 2008, Lange et al. 2010, Gibbs et al. 2012, Scheiermann et al. 2013, Labrecque and Cermakian 2015, Ella et al. 2016). Together, these molecular and cellular rhythms influence rhythmic immunological processes at the organismal level.

Mice show differential survival against infection in a time-of-day dependent manner. Some other immune related processes, such as inflammation, immune resistance, and the severity of autoimmune diseases, like rheumatoid arthritis, are found to vary throughout the day in a rhythmic manner (Cutolo 2012, Gibbs and Ray 2013, Curtis et al. 2014, Carter et al. 2016). The chronic disruption of circadian rhythms, such as sleep deprivation, shift work, and jet lag can precipitate disease even in healthy individuals and exacerbate existing diseases, particularly inflammatory conditions (Ranjbaran et al. 2007).

Drosophila melanogaster is a model organism that is widely used to study the mechanisms of humoral immune response (Figure 1.4). The humoral immune system in *Drosophila* is simplified, yet highly conserved at a molecular level with its mammalian counterpart (Muller et al. 2008). The discovery and understanding of mammalian pattern recognition receptors are triggered by the discovery of *Toll* in *Drosophila* immune system (Anderson 2000, Kimbrell and Beutler 2001,

Hoffmann 2003). As shown in Figure 1.4, there are two major pathways, *Toll* pathway and *Imd* (immune deficiency) pathway, within humoral immune system of *Drosophila*. *Toll* pathway mainly combats gram- positive bacterial infection and fungal infection, while *Imd* pathway mainly combat gram-negative bacterial infection.

There are pattern recognition proteins either circulating in the hemolymph or located on the cell membrane of the fat body for each pathway. These pattern recognition proteins could distinguish specific pathogen-associated molecular patterns and activate the downstream signaling pathway to induce the expression of AMPs (anti-microbial peptides) within the immune system, particularly within the fat body. AMPs are then secreted into the hemolymph to fight against the invading pathogen (Imler and Hoffmann 2000, Hoffmann 2003).

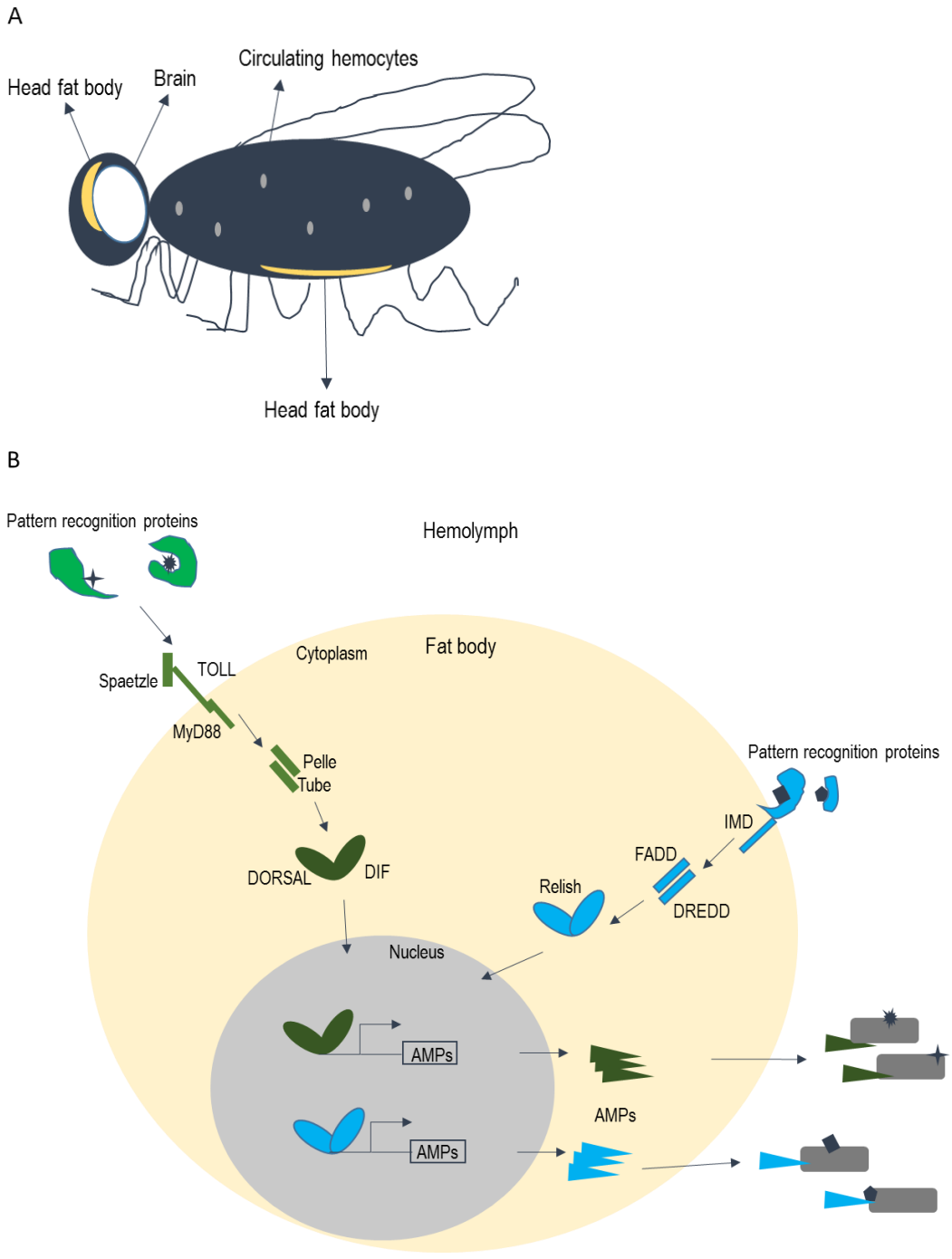
Similar to mammals, the immune response in *Drosophila* is rhythmic. Several genes involved in immune response are rhythmically expressed, and flies infected with pathogenic bacteria at different times of the day show a rhythmic resistance peaking during the late night (McDonald and Rosbash 2001, Lee and Edery 2008, Stone et al. 2012). However, it is unclear how this is regulated at either a molecular or cellular level.

Current questions and outline

As described above, one of questions waiting to be answered is, how does the principal oscillator regulates rhythmic physiology in peripheral tissues? We decide to choose the regulation of immune system in *Drosophila* as our model to study.

Drosophila has several unique advantages: First, flies are easy to raise. Flies grow fast, they can complete one life cycle in ten to twelve days. Flies are tiny, they are only a few millimeters in

Figure 1. 4. The immune system in *Drosophila*



A. Drosophila anatomy showing major immunological tissues: two fat bodies, one located in the

head and the other one in the abdomen. And the hemocytes circulating within the cavity.

B. A schematic presentation of the immune pathways*, Toll* pathway and *Imd* pathway in *Drosophila.* This figure is modified from Jules A Hoffmann, (2003) Nature (Hoffmann 2003).

body length. It is thus easy to maintain a large amount of flies in lab. Second, flies have a smaller genome, and there are various transgenic lines for most fly genes. Besides, there are well established genetic tools to get use of these transgenic lines. For example, UAS (upstream activating sequence)-*Gal4* system is an extensively used genetic tool to manipulate the expression of certain genes in a tissue specific manner (Brand and Perrimon 1993). Briefly, there is one parental fly line with a *Gal4* coding sequence downstream of a tissue specific promoter inserted into the genome, tissue specific GAL4 protein is expressed in this line. There is a second parental fly line with a transgenic fragment downstream of the UAS promoter inserted into the genome. Depending on the purpose, either a coding sequence or an RNAi fragment can be used as transgenic fragment here. By crossing these two parental lines, expressed GAL4 protein will bind to the UAS element and drive the expression of that transgenic fragment in the F1 offspring. Third, flies share conserved mechanism with mammals. Over half of human disease related genes have their counterparts in flies, and the regulation mechanisms of most processes are conserved from flies to mice. For example, as show in Figure 1.3, the architecture of the molecular clock is conserved in both flies and mice.

We then choose the immune system to study because 1), it is nicely rhythmic with a clear readout. We can infect flies with pathogenic bacterial and look at the survival. And we can monitor the expression of immune response genes as well as bacterial growth after infection in flies. 2), the clock in the fat body damps quickly under constant condition, suggesting that it relies largely on the principal oscillator to regulate its rhythmicity.

Using this *Drosophila* immune system as a working model, we found *Achilles (Achl),* a clock controlled gene that is expressed in the brain, regulates the immune system (Chapter 3) as well as its rhythmicity (Chapter 4). We also profiled candidate CCGs downstream of *Achl* using high

throughput RNA-seq (Chapter 4). To determine the best experimental conditions for using RNA-seq to profile circadian genes, we performed a computational simulation for various conditions based on published experimental data (Chapter 3).

Bibliography

Allada, R. and B. Y. Chung (2010). "Circadian organization of behavior and physiology in Drosophila." Annu Rev Physiol **72**: 605-624.

Anderson, K. V. (2000). "Toll signaling pathways in the innate immune response." Current Opinion in Immunology **12**(1): 13-19.

Balsalobre, A., L. Marcacci and U. Schibler (2000). "Multiple signaling pathways elicit circadian gene expression in cultured Rat-1 fibroblasts." Current Biology **10**(20): 1291-1294.

Brand, A. H. and N. Perrimon (1993). "Targeted gene expression as a means of altering cell fates and generating dominant phenotypes." Development **118**(2): 401-415.

Carter, S. J., H. J. Durrington, J. E. Gibbs, J. Blaikley, A. S. Loudon, D. W. Ray and I. Sabroe (2016). "A matter of time: study of circadian clocks and their role in inflammation." J Leukoc Biol **99**(4): 549-560.

Covington, M. F., J. N. Maloof, M. Straume, S. A. Kay and S. L. Harmer (2008). "Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development." Genome Biol **9**(8): R130.

Curtis, Anne M., Marina M. Bellet, P. Sassone-Corsi and Luke A. J. O'Neill (2014). "Circadian Clock Proteins and Immunity." Immunity **40**(2): 178-186.

Curtis, A. M., C. T. Fagundes, G. Yang, E. M. Palsson-McDermott, P. Wochal, A. F. McGettrick, N. H. Foley, J. O. Early, L. Chen, H. Zhang, C. Xue, S. S. Geiger, K. Hokamp, M. P. Reilly, A. N. Coogan, E. Vigorito, G. A. FitzGerald and L. A. J. O'Neill (2015). "Circadian control of innate immunity in macrophages by miR-155 targeting Bmal1." Proceedings of the National Academy of Sciences **112**(23): 7231-7236.

Cutolo, M. (2012). "Chronobiology and the treatment of rheumatoid arthritis." Curr Opin Rheumatol **24**(3): 312-318.

Dunlap, J. C. (1999). "Molecular Bases for Circadian Clocks." Cell **96**(2): 271-290.

Ella, K., R. Csepanyi-Komi and K. Kaldi (2016). "Circadian regulation of human peripheral neutrophils." Brain Behav Immun **57**: 209-221.

Filichkin, S. and T. Mockler (2012). "Unproductive alternative splicing and nonsense mRNAs: A widespread phenomenon among plant circadian clock genes." Biology Direct **7**(1): 1-15.

Gibbs, J. E., J. Blaikley, S. Beesley, L. Matthews, K. D. Simpson, S. H. Boyce, S. N. Farrow, K.

J. Else, D. Singh, D. W. Ray and A. S. I. Loudon (2012). "The nuclear receptor REV-ERBα mediates circadian regulation of innate immunity through selective regulation of inflammatory cytokines." Proceedings of the National Academy of Sciences **109**(2): 582-587.
Gibbs, J. E. and D. W. Ray (2013). "The role of the circadian clock in rheumatoid arthritis." Arthritis Res Ther **15**(1): 205.
Halberg, F., G. Cornelissen, W. Ulmer, M. Blank, W. Hrushesky, P. Wood, R. K. Singh and Z. Wang (2006). "Cancer chronomics III. Chronomics for cancer, aging, melatonin and experimental therapeutics researchers." J Exp Ther Oncol **6**(1): 73-84.
Harmer, S. L., J. B. Hogenesch, M. Straume, H.-S. Chang, B. Han, T. Zhu, X. Wang, J. A. Kreps and S. A. Kay (2000). "Orchestrated Transcription of Key Pathways in Arabidopsis by the Circadian Clock." Science **290**(5499): 2110-2113.
Hastings, M. H., A. B. Reddy and E. S. Maywood (2003). "A clockwork web: circadian timing in brain and periphery, in health and disease." Nat Rev Neurosci **4**(8): 649-661.
Hetru, C., L. Troxler and J. A. Hoffmann (2003). "Drosophila melanogaster antimicrobial defense." J Infect Dis **187 Suppl 2**: S327-334.
Hoffmann, J. A. (2003). "The immune response of Drosophila." Nature **426**(6962): 33-38.
Hsu, P. Y. and S. L. Harmer (2014). "Global profiling of the circadian transcriptome using microarrays." Methods Mol Biol **1158**: 45-56.
Hughes, M. E., G. R. Grant, C. Paquin, J. Qian and M. N. Nitabach (2012). "Deep sequencing the circadian and diurnal transcriptome of Drosophila brain." Genome Res **22**(7): 1266-1281.
Imler, J.-L. and J. A. Hoffmann (2000). "Signaling mechanisms in the antimicrobial host defense of Drosophila." Current Opinion in Microbiology **3**(1): 16-22.
Keegan, K. P., S. Pradhan, J. P. Wang and R. Allada (2007). "Meta-analysis of Drosophila circadian microarray studies identifies a novel set of rhythmically expressed genes." PLoS Comput Biol **3**(11): e208.
Keller, M., J. Mazuch, U. Abraham, G. D. Eom, E. D. Herzog, H. D. Volk, A. Kramer and B. Maier (2009). "A circadian clock in macrophages controls inflammatory immune responses." Proc Natl Acad Sci U S A **106**(50): 21407-21412.
Kimbrell, D. A. and B. Beutler (2001). "The evolution and genetics of innate immunity." Nat Rev Genet **2**(4): 256-267.
Klerman, E. B. (2005). "Clinical aspects of human circadian rhythms." J Biol Rhythms **20**(4): 375-386.

Ko, C. H. and J. S. Takahashi (2006). "Molecular components of the mammalian circadian clock." Hum Mol Genet **15 Spec No 2**: R271-277.

Kornmann, B., O. Schaad, H. Bujard, J. S. Takahashi and U. Schibler (2007). "System-driven and oscillator-dependent circadian transcription in mice with a conditionally active liver clock." PLoS Biol **5**(2): e34.

Labrecque, N. and N. Cermakian (2015). "Circadian Clocks in the Immune System." J Biol Rhythms **30**(4): 277-290.

Lange, T., S. Dimitrov and J. Born (2010). "Effects of sleep and circadian rhythm on the human immune system." Ann N Y Acad Sci **1193**: 48-59.

Lee, J.-E. and I. Edery (2008). "Circadian Regulation in the Ability of Drosophila to Combat Pathogenic Infections." Current Biology **18**(3): 195-199.

Levi, F. and U. Schibler (2007). "Circadian rhythms: mechanisms and therapeutic implications." Annu Rev Pharmacol Toxicol **47**: 593-628.

McDonald, M. J. and M. Rosbash (2001). "Microarray analysis and organization of circadian gene expression in Drosophila." Cell **107**(5): 567-578.

McGlincy, N. J., A. Valomon, J. E. Chesham, E. S. Maywood, M. H. Hastings and J. Ule (2012). "Regulation of alternative splicing by the circadian clock and food related cues." Genome Biol **13**(6): R54.

Mendez-Ferrer, S., D. Lucas, M. Battista and P. S. Frenette (2008). "Haematopoietic stem cell release is regulated by circadian oscillations." Nature **452**(7186): 442-447.

Muller, U., P. Vogel, G. Alber and G. A. Schaub (2008). "The innate immune system of mammals and insects." Contrib Microbiol **15**: 21-44.

Nagoshi, E., C. Saini, C. Bauer, T. Laroche, F. Naef and U. Schibler (2004). "Circadian gene expression in individual fibroblasts: cell-autonomous and self-sustained oscillators pass time to daughter cells." Cell **119**(5): 693-705.

Nitabach, M. N. and P. H. Taghert (2008). "Organization of the Drosophila circadian control circuit." Curr Biol **18**(2): R84-93.

O'Neill, J. S. and A. B. Reddy (2011). "Circadian clocks in human red blood cells." Nature **469**(7331): 498-503.

Ranjbaran, Z., L. Keefer, E. Stepanski, A. Farhadi and A. Keshavarzian (2007). "The relevance of sleep abnormalities to chronic inflammatory conditions." Inflamm Res **56**(2): 51-57.

Scheiermann, C., Y. Kunisaki and P. S. Frenette (2013). "Circadian control of the immune

system." Nat Rev Immunol **13**(3): 190-198.

Schmidt, T. M., M. T. H. Do, D. Dacey, R. Lucas, S. Hattar, Matynia and Anna (2011). "Melanopsin-positive Intrinsically Photosensitive Retinal Ganglion Cells: From Form to Function." The Journal of neuroscience : the official journal of the Society for Neuroscience **31**(45): 16094-16101.

Silver, A. C., A. Arjona, M. E. Hughes, M. N. Nitabach and E. Fikrig (2012a). "Circadian expression of clock genes in mouse macrophages, dendritic cells, and B cells." Brain, behavior, and immunity **26**(3): 407-413.

Silver, A. C., A. Arjona, W. E. Walker and E. Fikrig (2012b). "The circadian clock controls toll-like receptor 9-mediated innate and adaptive immunity." Immunity **36**(2): 251-261.

Slat, E., G. M. Freeman, Jr. and E. Herzog (2013). The Clock in the Brain: Neurons, Glia, and Networks in Daily Rhythms. Circadian Clocks. A. Kramer and M. Merrow, Springer Berlin Heidelberg. **217:** 105-123.

Stone, E. F., B. O. Fulton, J. S. Ayres, L. N. Pham, J. Ziauddin and M. M. Shirasu-Hiza (2012). "The circadian clock protein timeless regulates phagocytosis of bacteria in Drosophila." PLoS Pathog **8**(1): e1002445.

Trapnell, C., B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold and L. Pachter (2010). "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation." Nat Biotechnol **28**(5): 511-515.

Watson, F. L., R. Puttmann-Holgado, F. Thomas, D. L. Lamar, M. Hughes, M. Kondo, V. I. Rebel and D. Schmucker (2005). "Extensive diversity of Ig-superfamily proteins in the immune system of insects." Science **309**(5742): 1874-1878.

Wichert, S., K. Fokianos and K. Strimmer (2004). "Identifying periodically expressed transcripts in microarray time series data." Bioinformatics **20**(1): 5-20.

Wijnen, H., F. Naef and M. W. Young (2005). Molecular and Statistical Tools for Circadian Transcript Profiling. Methods in Enzymology. W. Y. Michael, Academic Press. **Volume 393:** 341-365.

Yoo, S. H., S. Yamazaki, P. L. Lowrey, K. Shimomura, C. H. Ko, E. D. Buhr, S. M. Siepka, H. K. Hong, W. J. Oh, O. J. Yoo, M. Menaker and J. S. Takahashi (2004). "PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues." Proc Natl Acad Sci U S A **101**(15): 5339-5346.

Zhang, R., N. F. Lahens, H. I. Ballance, M. E. Hughes and J. B. Hogenesch (2014). "A circadian gene expression atlas in mammals: Implications for biology and medicine." Proc Natl Acad Sci U S A.

Chapter 2: Considerations for RNA-sequencing analysis of circadian rhythms

Abstract

Circadian rhythms are daily endogenous oscillations of behavior, metabolism, and physiology. At a molecular level, these oscillations are generated by transcriptional–translational feedback loops composed of core clock genes. In turn, core clock genes drive the rhythmic accumulation of downstream outputs—termed clock-controlled genes (CCGs)—whose rhythmic translation and function ultimately underlie daily oscillations at a cellular and organismal level. Given the circadian clock's profound influence on human health and behavior, considerable efforts have been made to systematically identify CCGs. The recent development of next-generation sequencing has dramatically expanded our ability to study the expression, processing, and stability of rhythmically expressed mRNAs. Nevertheless, like any new technology, there are many technical issues to be addressed. Here, we discuss considerations for studying circadian rhythms using genome scale transcriptional profiling, with a particular emphasis on RNA sequencing. We make a number of practical recommendations—including the choice of sampling density, read depth, alignment algorithms, read-depth normalization, and cycling detection algorithms—based on computational simulations and our experience from previous studies. We believe that these results will be of interest to the circadian field and help investigators design experiments to derive most values from these large and complex data sets.

Introduction

Circadian rhythms are daily endogenous oscillations of behavior, physiology and metabolism that allow organisms to anticipate and respond to predictable environmental changes. In animals, these oscillations are governed by a dedicated timing system composed in large part by transcriptional-translational feedback loops of core clock genes (Ko and Takahashi 2006). At an

organismal level, circadian rhythms have profound influence over normal physiological rhythms such as sleep / wake cycles, while disruption of the clock contributes to many human disorders, including cardiovascular disease, neurodegenerative disease, obesity, diabetes, and cancer (Klerman 2005, Halberg et al. 2006, Levi and Schibler 2007).

In both mammals and insects, the principal circadian oscillator resides in a small number of neurons in the central nervous system (Nitabach and Taghert 2008, Slat et al. 2013). The molecular circadian clock in these neurons is entrained by external stimuli, ultimately synchronizing organismal rhythms. In mammals, this central clock is located in the suprachiasmatic nuclei (SCN) of the hypothalamus (Hastings et al. 2003, Stratmann and Schibler 2006). SCN neurons receive both photic and non-photic information from the environment and coordinate behavioral rhythms in locomotion, feeding, and sleep / wake cycles.

Through both direct and indirect mechanisms, the SCN also synchronizes downstream molecular circadian clocks in the brain and in peripheral tissues throughout the body. Peripheral clocks are typically phase-delayed from the SCN by 4-6 hours (Panda et al. 2002b) but otherwise have many of the same genetic and biochemical properties of clocks in the central oscillator. Notably, peripheral oscillations are endogenous and self-sustaining, persisting for days or even weeks *in vitro* (Yoo et al. 2004). Even cultured cell lines that have been maintained *in vitro* for many years maintain endogenous circadian oscillators that can be synchronized by a variety of stimuli (Balsalobre et al. 2000, Nagoshi et al. 2004). The discovery of circadian rhythms in tissue culture has had an enormous impact on the field, as these cellular circadian models have proven to be a fruitful resource for investigating core clock mechanisms (Baggs et al. 2009, Zhang et al. 2009).

In both central and peripheral oscillators, core clock proteins drive the rhythmic expression of

downstream targets, which are termed “clock controlled genes” (CCGs). These output genes do not participate directly in the mechanism of the circadian timekeeper, but instead are translated and ultimately impose rhythmicity on downstream cellular and physiological functions (Hastings et al. 2003). Many CCGs regulate the rate-limiting steps of metabolic and genetic pathways, indicating that they play a key role in temporally compartmentalizing cellular functions (Panda et al. 2002b). Although a systematic review of every CCG with an established molecular function is beyond the scope of this manuscript, it is worth emphasizing that maintaining appropriate rhythmic expression of single genes can be a matter of life and death. For example, a number of key ion channels are under circadian control in cardiomyocytes, and dysregulation of their rhythmic expression has profound consequences for the physiology of the heart while predisposing animals to fatal arrhythmia (Jeyaraj et al. 2012, Schroder et al. 2013).

The total number of cycling transcripts in any given tissue is difficult to ascertain and depends on many assumptions, but we can be confident that it ranges from a few hundred to several thousand transcripts, depending on the tissue (Hughes et al. 2007, Hughes et al. 2009). Notably, although the core clock machinery is largely conserved in different tissues, circadian output genes are highly tissue-specific (Ceriani et al. 2002, Panda et al. 2002a, Storch et al. 2002, Hughes et al. 2009). This observation makes intuitive sense, as the physiological demands on the liver, for example, are substantially different from those on neural tissues. But the diversity of CCGs complicates matters for investigators studying the molecular mechanisms of circadian clock output, and it provides strong motivation for experiments aimed at systematically identifying CCGs in different tissues and species.

To identify CCGs and understand the mechanism of their regulation, microarrays have been used to profile rhythmic gene expression systematically in cyanobacteria, plants, insects, fungi,

mice, and cellular models (Keegan et al. 2007, Kornmann et al. 2007, Menger et al. 2007, Covington et al. 2008, Hughes et al. 2009, Vollmers et al. 2009, Atwood et al. 2011, Rund et al. 2011, Xu et al. 2011, Hughes et al. 2012b, McGlincy et al. 2012). These studies have contributed significantly to our understanding of circadian output in *wildtype* animals, and over time they have matured into investigations of more focused tissues and cell types ((Kula-Eversole et al. 2010, Collins et al. 2012)). Most of these data sets are freely available online and provide a powerful resource for researchers interested in visualizing the expression of multiple genes in many human and mouse tissues (Pizarro et al. 2013, Zhang et al. 2014). Moreover, these data have contributed significantly to computational modeling studies of the molecular mechanism of circadian rhythms (Bozek et al. 2009, Anafi et al. 2014).

Besides simply cataloging CCGs, microarray profiling of rhythmic gene expression has also been instrumental in elucidating the mechanism of circadian output pathways. For example, a pair of recent studies used microarrays in conjunction with tissue-specific manipulation of clock genes to explore the relationship between central and peripheral oscillators (Kornmann et al. 2007, Hughes et al. 2012b). Both studies show that the peripheral circadian clock is essential for normal CCG expression and identified a number of candidate genes that may coordinate the synchronization of peripheral rhythms. A similar approach has characterized the fundamental role feeding cues have in driving CCG expression in peripheral tissues. By simply manipulating the time of day at which mice are allowed to feed, the phase of most CCGs in the liver was dramatically changed, underscoring the complexity of CCG regulation in the periphery (Vollmers et al. 2009). Finally, a recent study of rhythmic gene expression in *Dicer*-mutant fruit flies has shown the key role that miRNAs have in regulating circadian rhythms and transcriptional output (Kadener et al. 2009). The role of miRNAs in clock regulation has been confirmed in mouse

(Chen et al. 2013), thus motivating follow-up experiments aimed at understanding in much greater detail the interplay between ncRNAs and mRNAs in circadian output.

The development of next generation sequencing (NGS) has accelerated studies into the global regulation of gene expression, and these technical advances offer significant opportunities for the circadian field. Early studies using RNA-sequencing (RNA-seq) to profile circadian gene expression have demonstrated the potential of these approaches (Filichkin and Mockler 2012, Hughes et al. 2012a, Menet et al. 2012, Du et al. 2014). Moreover, the single base-pair resolution of these data enables the detection of new cycling transcripts, as well as measuring alternative splice forms, RNA editing, and other forms of RNA processing. Besides RNA-seq, there are other emerging NGS technologies that have begun to influence circadian research. For example, nascent RNA-seq (Menet et al. 2012, Rodriguez et al. 2013) provides information about the transcriptional and post-transcriptional regulation of cycling mRNAs. ChIP-seq enables the characterization of how these output rhythms are regulated by transcription factor binding and chromatin regulation (Rey et al. 2011, Bugge et al. 2012, Koike et al. 2012, Meireles-Filho et al. 2014, Menet et al. 2014).

Despite the impact that NGS has had on circadian research, many technical challenges await investigators conducting these experiments. Several of these challenges involving experimental and statistical design are common to all global gene expression studies, including those using microarrays. We point the interested reader to several excellent articles that have discussed these issues in detail (Walker and Hogenesch 2005, Wijnen et al. 2005, Deckard et al. 2013, Hsu and Harmer 2014). In addition, the use of RNA-seq introduces a number of technical issues that have not been satisfactorily addressed by the circadian field, such as the depth of sequencing coverage, read-depth normalization, and choice of alignment algorithm. Here, we

present recommendations for future work using RNA-seq to explore circadian mRNA rhythms, with a focus on the computational and statistical approaches necessary for data interpretation.

Conceptually, the systematic identification of CCGs is remarkably straight-forward. Tissue samples are collected at regular intervals, mRNA expression is measured on a global scale for each of these samples, and appropriate statistical tests are used to identify rhythmic components of the data. Typically, tissue collections are performed in constant darkness in order to isolate rhythms driven by the circadian clock, but many valuable studies have also been performed under LD conditions or in the presence of different zeitgebers. The key considerations when designing circadian RNA-seq experiments include: (1) number of time points and replicates, (2) choice of alignment algorithm, (3) method of read-depth normalization, (4) number of reads per sample, and (5) choice of statistical analyses and interpretation of the results.

Sample density

Statistical tests for rhythmicity are extremely sensitive to the frequency of sampling (Hughes et al. 2007, Hughes et al. 2009, Atwood et al. 2011). However, given the expense and relative novelty of RNA-seq, there are presently no circadian RNA-seq studies using very dense sampling schemes. To simulate higher sampling densities using the available data, we randomly combined subsets of reads from neighboring time points. These synthetic data points were then used to calculate expression levels at intermediate data points and thus gain a measure of the relationship between sampling density and cycling identification. We emphasize that this computational approach is an expedient to generate synthetic test data, rather than an approach to identify bona fide cycling transcripts. Nevertheless, we note that over half of the top cycling transcripts in these data also cycle in previous microarray studies (Hughes et al. 2009, Hughes

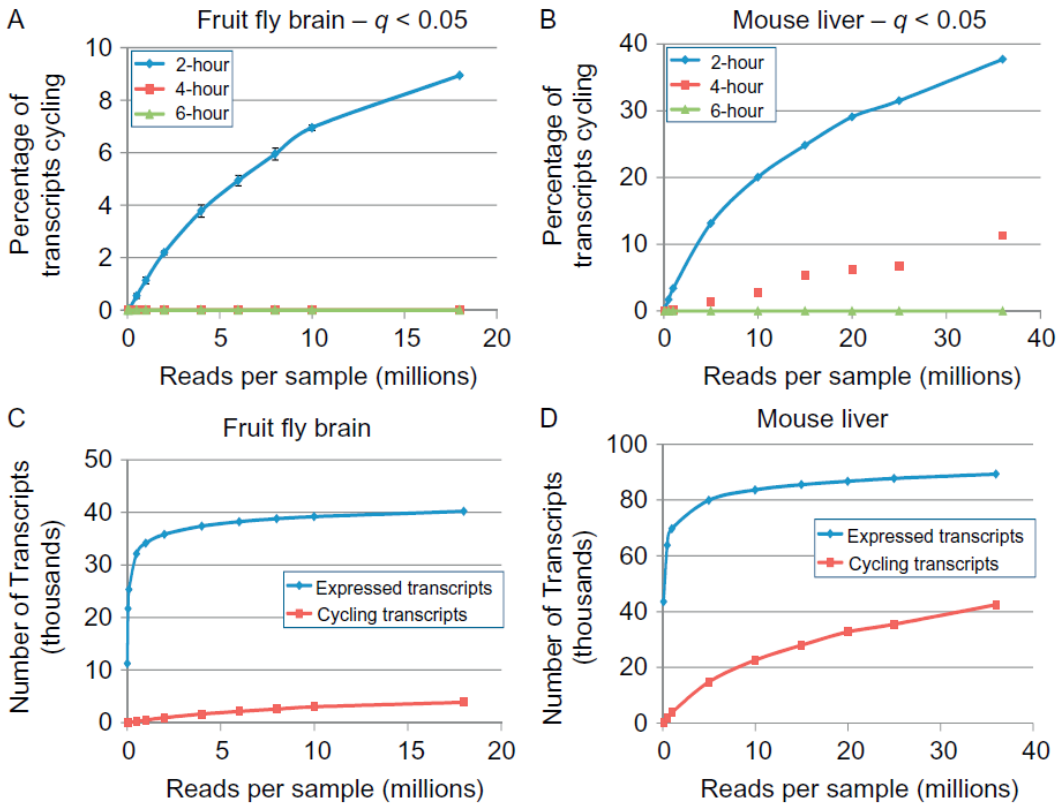
et al. 2012a). The total number of cycling transcripts and their distribution of phases and amplitudes in these data were also consistent with previous studies. Most importantly, the number of uniquely aligned reads and the dynamic range of transcript expression were both within normal ranges, indicating that our synthetic data realistically model the properties of a circadian RNA-seq experiment.

We found that 2-h sampling resolution over two consecutive days dramatically increases the number of identified cycling transcripts relative to 4- and 6-h sampling schemes (Figure 2.1A and 2.1B). Moreover, the identification of cycling transcripts at 2-h resolution yielded considerably fewer false positives, which we determined by comparing to cycling transcripts identified in period-null fruit flies (data not shown). These results agree with the previous circadian microarray studies mentioned above, and based on their consensus, we strongly recommend using 2-h sampling over two consecutive days. The approximately twofold increase in cost is more than compensated by dramatic improvements in the accuracy and reproducibility of cycling identification.

Alignment algorithm and splice form detection

Many different algorithms have been developed to align raw RNA-seq reads to a reference genome/transcriptome and calculate expression values. These algorithms, including but not limited to Bowtie, Tophat, RUM, Star, and GSNAP have unique strengths and weaknesses based on speed, memory footprint, sensitivity, and accuracy(Langmead et al. 2009, Trapnell et al. 2009, Trapnell et al. 2010, Wu and Nacu 2010, Grant et al. 2011, Dobin et al. 2012, Kim et al. 2013). We recommend RUM because it is robust, user- friendly, and exceptionally good at mapping reads to exon–exon junctions. However, we note that other algorithms each have

Figure 2.1. The discovery of cycling transcripts depends on sampling density and read depth



To assess the relationship among sampling density, read depth, and the identification of cycling transcripts, subsets of raw reads were randomly selected from legacy data sets and used to measure gene expression. Two-hour sampling resolution was simulated from these data by randomly pooling reads from neighboring time points. The discovery cycling transcripts in the fruit fly brain (A) and the mouse liver (B) showed a clear positive dependence on total read depth and sampling density. The total number of expressed transcripts (>10 uniquely aligned reads across the entire data set) is plotted as a function of read depth per sample for fruit fly brain (C)

and mouse liver (D). Note that the blue traces (dark gray in the print version) in A and B have been replotted in C and D for the sake of clarity. In both data sets, the total number of cycling transcripts does not plateau, even at maximum read depths. Similarly, although the total number of expressed transcripts begins to plateau (2.5 million reads per sample for flies; 5 million reads per sample for mice), expressed transcripts continue to be identified even at maximal read depths.

specific advantages. Star, for example, is particularly effective when working with large data sets, being orders of magnitude faster than RUM, GSNAP, or Tophat.

A related consideration is splice form detection. Several methods have been proposed to identify and quantify splice forms, such as Cufflinks, Scripture, CEM, and iReckon (Trapnell et al. 2010). In (Hayer et al. 2015), using BEERS, we simulated up to 10 forms of 5000 Refseq genes. When detecting one or two forms, most algorithms are able to detect with reasonable accuracy the internal gene structures of these models. However, when three or more forms are included, both the false discovery and false negative rates become substantial. Put simply, when you need these algorithms to detect multiple forms, they fail more often than not, even using 100 bp paired-end reads. Because of their ineffectiveness in this regard, it is difficult to evaluate their quantification properties. While increasing read length may improve splice form detection (250 bp paired-end reads are now possible), at this point, this is an unresolved problem. Alternatively, cycling analysis may be performed at an exon-level to thereby avoid the difficulties of accurately detecting alternative splice forms.

Read-depth normalization

Read depth determines signal to noise in detection and, consequently, the ability to detect cycling. Unfortunately, with RNA-seq it is impossible to get the same number of reads for multiple samples. This is in contrast to arrays, where normalization methods such as RMA and GCRMA are robust for small changes in overall signal between arrays. If the number of reads per sample is roughly equal, downsampling can be an attractive option. In this way, you “fix” the read depth to the sample for which you have the fewest mapped reads. For example, in the above fly samples with 15–21 m reads, we could sample 15 m reads from each time point. This

has the unattractive property of throwing away data, but it is probably the best strategy to normalize read depth between samples.

Read depth

Read depth is a key factor that determines the accuracy of measurements made by high throughput sequencing (Hart et al. 2013, Liu et al. 2013, Jung et al. 2014). More reads increase the statistical power of gene expression measurements and allow the detection of rare transcripts. But unlike microarrays whose cost is driven by the number of samples, the cost of RNA-seq is largely determined by the number of sequenced reads. Therefore, care must be taken to avoid under-powered experiments on one hand and wasted resources on the other.

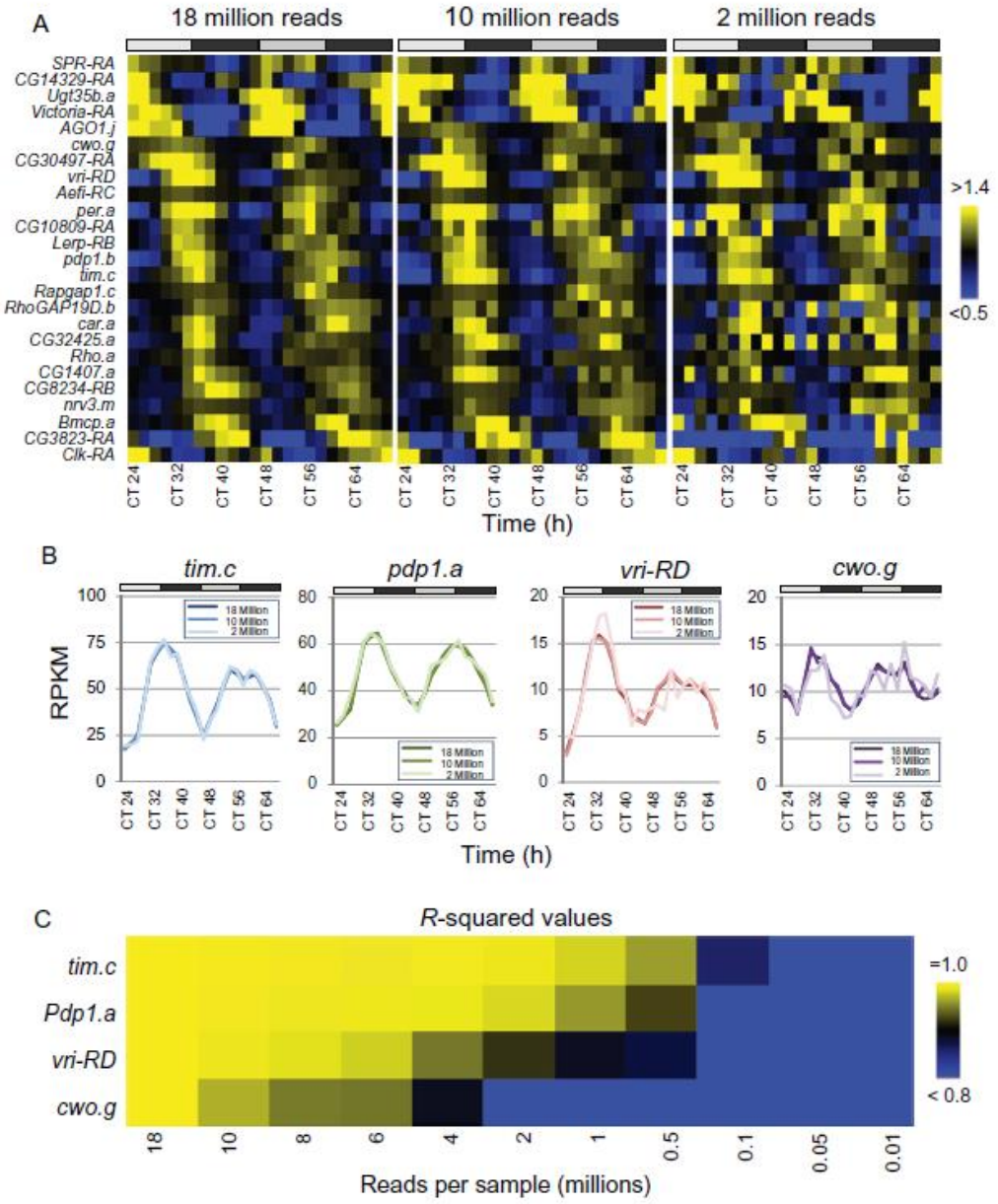
As seen in Figure 2.1, the number of identified cycling transcripts depends on read depth. Notably, the number of cyclers does not plateau, even at the maximum read depths available in this study. For comparison, we plotted the total number of detectable transcripts in these data, defined as 10 uniquely aligned reads across the entire data set (Figure 2.1C and 2.1D). Unlike cycling transcripts, expressed transcripts begin to plateau in these data, but some expressed transcripts are only detectable at very high read depths. Consistent with this, our unpublished data indicate that even one billion reads per sample may be insufficient to detect every expressed transcript in mouse tissues. It stands to reason that determining whether a transcript cycles requires considerably more reads than merely detecting its expression, and we speculate that hundreds of millions of reads per sample (at 2-h resolution or greater) will be required to identify every cycling transcript in any given tissue.

On the other hand, if the goal of an experiment is to assess changes in the circadian transcriptome rather than catalog every cycling transcript, far fewer reads per sample may be

necessary. This distinction is akin to the difference between de novo sequencing of a genome versus resequencing to identify allelic variants. To explore this relationship, we compared the expression profiles of 25 known cycling transcripts in the fly brain at different read depths (Figure 2.2A). We found that the overall rhythmic pattern of these transcripts is maintained at 50% the maximal read depth, and much of the rhythmic signal persists even at 10% of the maximal read depth. To illustrate this observation, we plotted the expression pattern of four clock genes, timeless, vrille, Pdp1, and cwo, whose rhythmicity is evident even at low read depths (Figure 2.2B). In fact, the correlation between downsampled expression profiles and the "true" expression profile is maintained at surprisingly low read depths. For example, as little as 500 thousand reads per sample are sufficient to detect rhythms in timeless and Pdp1 (Figure 2.2C).

We broadened this observation to the whole transcriptome by determining how many cycling transcripts (defined as cycling at maximum reads depths) are also identified at lower read depths (Figure 2.3). As one might expect, in both fruit flies and mice, high-amplitude cyclers were detectable at relatively low read depths. Similarly, highly expressed transcripts are more likely to be identified as cycling at lower read depths (data not shown). We used these observations to calculate a rough estimate of the number of reads per sample necessary to identify 50%, 75%, or 87.5% of the total circadian transcriptome (Figure 2.4). Since the maximal read depths available in this study are insufficient to identify every cycling transcript as discussed in detail above, we caution the reader that these figures represent low-end estimates. With that caveat in mind, however, these data indicate that 10–15 million reads per sample in flies and 20–25 million reads per sample in mice may be sufficient to characterize the majority of the circadian transcriptome, especially if the investigator is content to measure the highest amplitude and most highly expressed rhythmic genes.

Figure 2.2. Many cycling transcripts are detectable with far fewer reads per sample than seen in legacy data sets

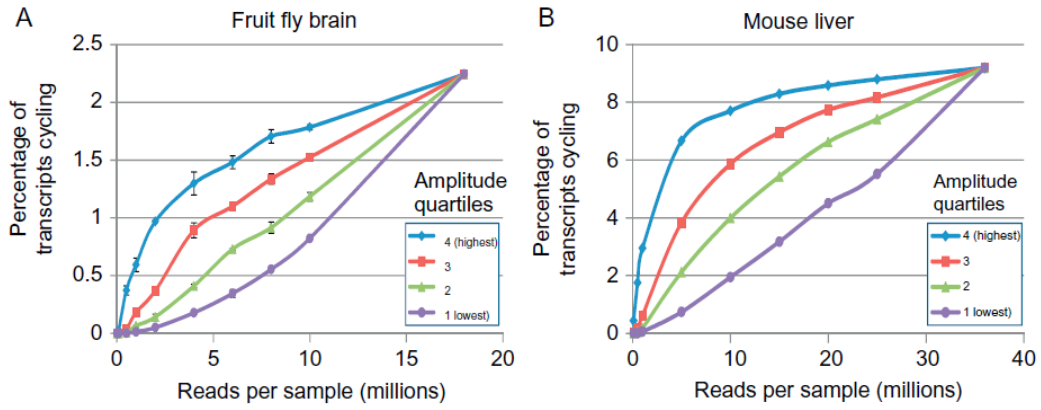


(A) The expression pattern of 25 cycling transcripts in the fruit fly brain at different read depths

are shown as heat maps. The top gray and black bars represent subjective day and night, respectively. Every transcript's expression profile has been median-normalized. Expression levels greater than 1.4-fold are shown as bright yellow; expression levels less than 0.5-fold are shown as bright blue (legend on the far right). (B) The rhythmic expression pattern of four representative transcripts,

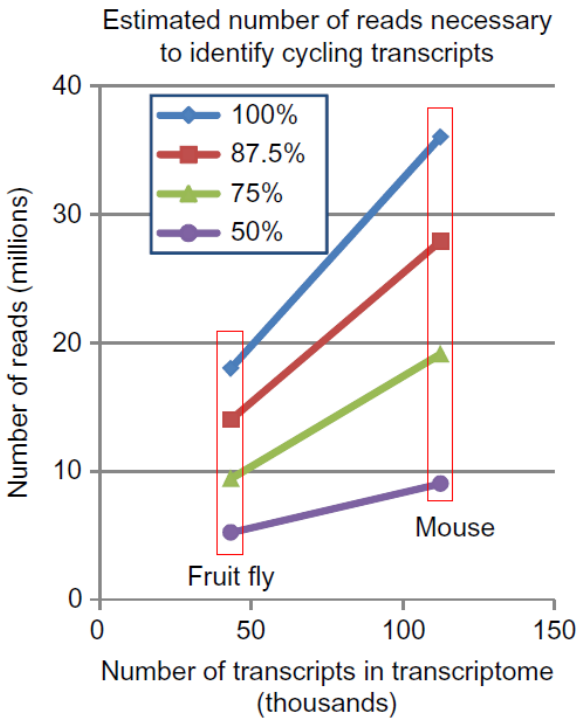
timeless.c, pdp1.a, vrille-RD, and *cwo.g* are shown at different read depths. The top gray and black bars represent subjective day and night, respectively. The vertical axis represents PKM values; the horizontal axis indicates time points which are marked on the bottom. (C) Pearson correlation coefficients between maximal read depth and subsampled data of expression values of the four transcripts in (B). The horizontal axis indicates different read depth. Correlations equal to 1.0 are shown as bright yellow; correlations less than 0.8 are shown as bright blue (legend on the far right).

Figure 2.3. Detection of cycling transcripts at lower read depth depends on amplitude



Cycling transcripts were divided in quartiles based on the magnitude of the amplitude. The percentage of cycling transcripts from each quartile is plotted as function of read depth in fruit fly brain (A) and mouse liver (B) data sets. Transcripts with higher amplitudes are more likely to be detected at a lower read depth, while the detection of low-amplitude transcripts requires increasingly large read depths. The vertical axis refers to the percentage of transcripts cycling (e.g., since 8% of transcripts in the fruit fly brain cycle, each quartile contains 2% of the transcriptome cycling, at max).

Figure 2.4. Estimated number of reads necessary to detect subsets of the circadian transcriptome



The number of reads per sample necessary to detect 50%, 75%, 87.5%, and 100% of cycling transcripts in the fruit fly and mouse transcriptome are shown here. The vertical axis indicates read depth per sample, and the horizontal axis indicates the total number of transcripts within the transcriptomes of fruit fly and mice.

Cycling detection algorithms

Many statistical tests have been developed to detect cycling components of complex data sets, including but not limited to COSOPT, Fisher's G test, F24, Haystack, Arser, JTK_Cycle, Lomb Scargle and Delichtenberg (Claridge-Chang et al. 2001, Straume 2004, Wichert et al. 2004, de Lichtenberg et al. 2005, Glynn et al. 2006, Mockler et al. 2007, Hughes et al. 2010, Yang and Su 2010). They are based on different mathematical assumptions, and their relative strengths and weaknesses have been compared in several recent benchmarking studies (Hughes et al. 2010, Deckard et al. 2013). In our on-going experiments, we favor using JTK_Cycle as it has proven to be powerful, accurate, and efficient, while also being relatively user-friendly. But we note that many of the related approaches have been used to great effect in previous cycling studies.

False discovery correction

Accurately estimating the rate of false discoveries is a well-recognized problem in any high-throughput analysis (Macarthur, 2012). For circadian experiments, the p-value is calculated independently for every transcript, and it reflects the probability that a given expression profile occurred by chance alone. However, this measure does not account for the enormous size of genomics experiments, which may include tens of thousands or even hundreds of thousands of expression profiles. Consequently, even extremely unlikely occurrences (i.e., transcripts with low p-values) are expected to be present with some frequency. Therefore, the false discovery rate (FDR or q-value) is calculated for the experiment as a whole, and it reflects the probability that transcripts at a given statistical threshold are in actuality false discoveries.

The rank order of p- and q-values will always be the same, and the q-value will inevitably be greater than or equal to the p-value (i.e., more conservative). Although the choice of statistical

threshold depends on the aims of the investigation and the tolerance for false discoveries, we strongly recommend the inclusion of an explicit false discovery correction to account for the enormous number of comparisons being made.

Validation and follow-up

Like any high-throughput assay, several steps should be taken to validate the accuracy gene expression measurements. Fortunately, circadian profiling experiments have excellent internal controls built-in, as the expression pattern of many core clock genes are known or predicted from previous work. Moreover, independent biological samples should be collected, and novel candidate cyclers should be verified using independent technical approaches such as quantitative PCR or in situ hybridization. Finally, comparisons with legacy data sets—both expression profiling and ChIP-seq—can be used to effectively strengthen the confidence that a novel cycling gene is worth pursuing in downstream functional experiments.

Discussion

Circadian rhythms exert an enormous influence on normal and pathological physiology. Nevertheless, the pathways by which the core circadian oscillator drives rhythms in peripheral tissues and the mechanisms through which peripheral clocks influence human health are incompletely understood. RNA-seq analysis of the circadian transcriptome can be used to build a comprehensive list of candidate CCGs as well as explore the underlying molecular biology through which these rhythms are generated. We expect that both applications will have a major impact on the circadian field in the years ahead. Care must be taken to use appropriate experimental designs and statistical approaches, particularly with respect to the sampling density and the number of reads per sample. Using experimental designs appropriate to the

goals of an investigation will ultimately increase the value and staying power of the data collected.

Materials and Methods

Data processing

Previously published circadian RNA-seq data from fruit fly brain (Hughes, Grant, et al., 2012) were downloaded from GEO (GSE36108). Mouse liver RNA-seq data were obtained in advance of publication, the GEO accession number is GSE54652 (Zhang et al. 2014). To simulate 2-h sampling in both data sets, randomly selected reads from neighboring time points were merged together using custom-built Ruby scripts; three independent replicates were generated for each data set in the fruit fly data.

Alignment

RUM (Grant et al., 2011) was used to align all reads (75 bp, paired-end) to the genome and transcriptome of either D. melanogaster (build dm3) or M. musculus (build mm9), using the following parameters: "–bowtie-nu-limit 10 –nu-limit 10." For the fruit fly data set, 52–59% of reads were mapped uniquely to the genome and transcriptome. For mouse data sets, 77–82% of reads were mapped uniquely to the genome and transcriptome. Reads per kilobase per million reads (RPKMs) for each transcript were calculated by RUM from uniquely mapped reads.

Circadian analysis: Detection of cycling was performed using either JTK_Cycle or JTK_Cycle_v2 (Hughes et al., 2010; Miyazaki et al., 2011) implemented in R (64-bit, version 2.12.1). Benjamini–Hochberg corrected q-values of<0.05 were generally used as statistical threshold. To mitigate the effect of false discovery in the fruit fly data set, *per0* data (i.e., circadian mutant)

were used as a negative control. Cycling transcripts found in *per0* samples were considered to be false discoveries. In the analysis of RPKM or amplitude, all the cycling transcripts in legacy data set were divided into four groups according to either mean RPKM or amplitude values, with the same numbers of transcripts in each group.

RUM can be downloaded and installed from the following: http://www.cbil.upenn.edu/RUM/userguide.php. JTK_Cycle is available from:

http://openwetware.org/wiki/HughesLab:JTK_Cycle. All Ruby scripts and data sets are available on demand.

Bibliography

Anafi, R. C., Y. Lee, T. K. Sato, A. Venkataraman, C. Ramanathan, I. H. Kavakli, M. E. Hughes, J. E. Baggs, J. Growe, A. C. Liu, J. Kim and J. B. Hogenesch (2014). "Machine Learning Helps Identify CHRONO as a Circadian Clock Component." PLoS Biol **12**(4): e1001840.

Atwood, A., R. DeConde, S. S. Wang, T. C. Mockler, J. S. M. Sabir, T. Ideker and S. A. Kay (2011). "Cell-autonomous circadian clock of hepatocytes drives rhythms in transcription and polyamine synthesis." Proceedings of the National Academy of Sciences **108**(45): 18560-18565.

Baggs, J. E., T. S. Price, L. DiTacchio, S. Panda, G. A. Fitzgerald and J. B. Hogenesch (2009). "Network features of the mammalian circadian clock." PLoS Biol **7**(3): e52.

Balsalobre, A., L. Marcacci and U. Schibler (2000). "Multiple signaling pathways elicit circadian gene expression in cultured Rat-1 fibroblasts." Current Biology **10**(20): 1291-1294.

Bozek, K., A. Relogio, S. M. Kielbasa, M. Heine, C. Dame, A. Kramer and H. Herzel (2009). "Regulation of clock-controlled genes in mammals." PLoS One **4**(3): e4882.

Bugge, A., D. Feng, L. J. Everett, E. R. Briggs, S. E. Mullican, F. Wang, J. Jager and M. A. Lazar (2012). "Rev-erbα and Rev-erbβ coordinately protect the circadian clock and normal metabolic function." Genes & Development **26**(7): 657-667.

Ceriani, M. F., J. B. Hogenesch, M. Yanovsky, S. Panda, M. Straume and S. A. Kay (2002). "Genome-wide expression analysis in Drosophila reveals genes controlling circadian behavior." J Neurosci **22**(21): 9305-9319.

Chen, R., M. D'Alessandro and C. Lee (2013). "miRNAs Are Required for Generating a Time Delay Critical for the Circadian Oscillator." Current Biology **23**(20): 1959-1968.

Claridge-Chang, A., H. Wijnen, F. Naef, C. Boothroyd, N. Rajewsky and M. W. Young (2001). "Circadian Regulation of Gene Expression Systems in the Drosophila Head." Neuron **32**(4): 657-671.

Collins, B., E. A. Kane, D. C. Reeves, M. H. Akabas and J. Blau (2012). "Balance of activity between LN(v)s and glutamatergic dorsal clock neurons promotes robust circadian rhythms in Drosophila." Neuron **74**(4): 706-718.

Covington, M. F., J. N. Maloof, M. Straume, S. A. Kay and S. L. Harmer (2008). "Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development." Genome Biol **9**(8): R130.

de Lichtenberg, U., L. J. Jensen, A. Fausboll, T. S. Jensen, P. Bork and S. Brunak (2005).

"Comparison of computational methods for the identification of cell cycle-regulated genes." Bioinformatics **21**(7): 1164-1171.

Deckard, A., R. C. Anafi, J. B. Hogenesch, S. B. Haase and J. Harer (2013). "Design and analysis of large-scale biological rhythm studies: a comparison of algorithms for detecting periodic signals in biological data." Bioinformatics **29**(24): 3174-3180.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson and T. R. Gingeras (2012). "STAR: ultrafast universal RNA-seq aligner." Bioinformatics **29**(1): 15-21.

Du, N. H., A. B. Arpat, M. De Matos and D. Gatfield (2014). "MicroRNAs shape circadian hepatic gene expression on a transcriptome-wide scale." Elife **3**: e02510.

Filichkin, S. and T. Mockler (2012). "Unproductive alternative splicing and nonsense mRNAs: A widespread phenomenon among plant circadian clock genes." Biology Direct **7**(1): 1-15.

Glynn, E. F., J. Chen and A. R. Mushegian (2006). "Detecting periodic patterns in unevenly spaced gene expression time series using Lomb-Scargle periodograms." Bioinformatics **22**(3): 310-316.

Grant, G. R., M. H. Farkas, A. D. Pizarro, N. F. Lahens, J. Schug, B. P. Brunk, C. J. Stoeckert, J. B. Hogenesch and E. A. Pierce (2011). "Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM)." Bioinformatics **27**(18): 2518-2528.

Halberg, F., G. Cornelissen, W. Ulmer, M. Blank, W. Hrushesky, P. Wood, R. K. Singh and Z. Wang (2006). "Cancer chronomics III. Chronomics for cancer, aging, melatonin and experimental therapeutics researchers." J Exp Ther Oncol **6**(1): 73-84.

Hart, S. N., T. M. Therneau, Y. Zhang, G. A. Poland and J. P. Kocher (2013). "Calculating sample size estimates for RNA sequencing data." J Comput Biol **20**(12): 970-978.

Hastings, M. H., A. B. Reddy and E. S. Maywood (2003). "A clockwork web: circadian timing in brain and periphery, in health and disease." Nat Rev Neurosci **4**(8): 649-661.

Hayer, K. E., A. Pizarro, N. F. Lahens, J. B. Hogenesch and G. R. Grant (2015). "Benchmark analysis of algorithms for determining and quantifying full-length mRNA splice forms from RNA-seq data." Bioinformatics **31**(24): 3938-3945.

Hsu, P. Y. and S. L. Harmer (2014). "Global profiling of the circadian transcriptome using microarrays." Methods Mol Biol **1158**: 45-56.

Hughes, M., L. Deharo, S. R. Pulivarthy, J. Gu, K. Hayes, S. Panda and J. B. Hogenesch (2007). "High-resolution time course analysis of gene expression from pituitary." Cold Spring Harb Symp Quant Biol **72**: 381-386.

Hughes, M. E., L. DiTacchio, K. R. Hayes, C. Vollmers, S. Pulivarthy, J. E. Baggs, S. Panda and J. B. Hogenesch (2009). "Harmonics of circadian gene transcription in mammals." PLoS Genet **5**(4): e1000442.

Hughes, M. E., G. R. Grant, C. Paquin, J. Qian and M. N. Nitabach (2012a). "Deep sequencing the circadian and diurnal transcriptome of Drosophila brain." Genome Res **22**(7): 1266-1281.

Hughes, M. E., J. B. Hogenesch and K. Kornacker (2010). "JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets." J Biol Rhythms **25**(5): 372-380.

Hughes, M. E., H. K. Hong, J. L. Chong, A. A. Indacochea, S. S. Lee, M. Han, J. S. Takahashi and J. B. Hogenesch (2012b). "Brain-specific rescue of Clock reveals system-driven transcriptional rhythms in peripheral tissue." PLoS Genet **8**(7): e1002835.

Jeyaraj, D., S. M. Haldar, X. Wan, M. D. McCauley, J. A. Ripperger, K. Hu, Y. Lu, B. L. Eapen, N. Sharma, E. Ficker, M. J. Cutler, J. Gulick, A. Sanbe, J. Robbins, S. Demolombe, R. V. Kondratov, S. A. Shea, U. Albrecht, X. H. Wehrens, D. S. Rosenbaum and M. K. Jain (2012). "Circadian rhythms govern cardiac repolarization and arrhythmogenesis." Nature **483**(7387): 96-99.

Jung, Y. L., L. J. Luquette, J. W. Ho, F. Ferrari, M. Tolstorukov, A. Minoda, R. Issner, C. B. Epstein, G. H. Karpen, M. I. Kuroda and P. J. Park (2014). "Impact of sequencing depth in ChIP-seq experiments." Nucleic Acids Res **42**(9): e74.

Kadener, S., J. S. Menet, K. Sugino, M. D. Horwich, U. Weissbein, P. Nawathean, V. V. Vagin, P. D. Zamore, S. B. Nelson and M. Rosbash (2009). "A role for microRNAs in the Drosophila circadian clock." Genes Dev **23**(18): 2179-2191.

Keegan, K. P., S. Pradhan, J. P. Wang and R. Allada (2007). "Meta-analysis of Drosophila circadian microarray studies identifies a novel set of rhythmically expressed genes." PLoS Comput Biol **3**(11): e208.

Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley and S. Salzberg (2013). "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions." Genome Biology **14**(4): R36.

Klerman, E. B. (2005). "Clinical aspects of human circadian rhythms." J Biol Rhythms **20**(4): 375-386.

Ko, C. H. and J. S. Takahashi (2006). "Molecular components of the mammalian circadian clock." Hum Mol Genet **15 Spec No 2**: R271-277.

Koike, N., S.-H. Yoo, H.-C. Huang, V. Kumar, C. Lee, T.-K. Kim and J. S. Takahashi (2012). "Transcriptional Architecture and Chromatin Landscape of the Core Circadian Clock in Mammals." Science **338**(6105): 349-354.

Kornmann, B., O. Schaad, H. Bujard, J. S. Takahashi and U. Schibler (2007). "System-driven and oscillator-dependent circadian transcription in mice with a conditionally active liver clock." PLoS Biol **5**(2): e34.

Kula-Eversole, E., E. Nagoshi, Y. Shang, J. Rodriguez, R. Allada and M. Rosbash (2010). "Surprising gene expression patterns within and between PDF-containing circadian neurons in Drosophila." Proceedings of the National Academy of Sciences **107**(30): 13497-13502.

Langmead, B., C. Trapnell, M. Pop and S. L. Salzberg (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." Genome Biol **10**(3): R25.

Levi, F. and U. Schibler (2007). "Circadian rhythms: mechanisms and therapeutic implications." Annu Rev Pharmacol Toxicol **47**: 593-628.

Liu, Y., J. F. Ferguson, C. Xue, I. M. Silverman, B. Gregory, M. P. Reilly and M. Li (2013). "Evaluating the impact of sequencing depth on transcriptome profiling in human adipose." PLoS One **8**(6): e66883.

McGlincy, N. J., A. Valomon, J. E. Chesham, E. S. Maywood, M. H. Hastings and J. Ule (2012). "Regulation of alternative splicing by the circadian clock and food related cues." Genome Biol **13**(6): R54.

Meireles-Filho, A. C., A. F. Bardet, J. O. Yanez-Cuna, G. Stampfel and A. Stark (2014). "cis-regulatory requirements for tissue-specific programs of the circadian clock." Curr Biol **24**(1): 1-10.

Menet, J. S., S. Pescatore and M. Rosbash (2014). "CLOCK:BMAL1 is a pioneer-like transcription factor." Genes Dev **28**(1): 8-13.

Menet, J. S., J. Rodriguez, K. C. Abruzzi and M. Rosbash (2012). "Nascent-Seq reveals novel features of mouse circadian transcriptional regulation." Elife **1**: e00011.

Menger, G. J., G. C. Allen, N. Neuendorff, S.-S. Nahm, T. L. Thomas, V. M. Cassone and D. J. Earnest (2007). Circadian profiling of the transcriptome in NIH/3T3 fibroblasts: comparison with rhythmic gene expression in SCN2.2 cells and the rat SCN.

Mockler, T. C., T. P. Michael, H. D. Priest, R. Shen, C. M. Sullivan, S. A. Givan, C. McEntee, S. A. Kay and J. Chory (2007). "The DIURNAL project: DIURNAL and circadian expression profiling, model-based pattern matching, and promoter analysis." Cold Spring Harb Symp Quant

Biol **72**: 353-363.

Nagoshi, E., C. Saini, C. Bauer, T. Laroche, F. Naef and U. Schibler (2004). "Circadian gene expression in individual fibroblasts: cell-autonomous and self-sustained oscillators pass time to daughter cells." Cell **119**(5): 693-705.

Nitabach, M. N. and P. H. Taghert (2008). "Organization of the Drosophila circadian control circuit." Curr Biol **18**(2): R84-93.

Panda, S., M. P. Antoch, B. H. Miller, A. I. Su, A. B. Schook, M. Straume, P. G. Schultz, S. A. Kay, J. S. Takahashi and J. B. Hogenesch (2002a). "Coordinated Transcription of Key Pathways in the Mouse by the Circadian Clock." Cell **109**(3): 307-320.

Panda, S., M. P. Antoch, B. H. Miller, A. I. Su, A. B. Schook, M. Straume, P. G. Schultz, S. A. Kay, J. S. Takahashi and J. B. Hogenesch (2002b). "Coordinated transcription of key pathways in the mouse by the circadian clock." Cell **109**(3): 307-320.

Pizarro, A., K. Hayer, N. F. Lahens and J. B. Hogenesch (2013). "CircaDB: a database of mammalian circadian gene expression profiles." Nucleic Acids Res **41**(Database issue): D1009-1013.

Rey, G., F. Cesbron, J. Rougemont, H. Reinke, M. Brunner and F. Naef (2011). "Genome-wide and phase-specific DNA-binding rhythms of BMAL1 control circadian output functions in mouse liver." PLoS Biol **9**(2): e1000595.

Rodriguez, J., C.-H. A. Tang, Y. L. Khodor, S. Vodala, J. S. Menet and M. Rosbash (2013). "Nascent-Seq analysis of Drosophila cycling gene expression." Proceedings of the National Academy of Sciences **110**(4): 275-284.

Rund, S. S. C., T. Y. Hou, S. M. Ward, F. H. Collins and G. E. Duffield (2011). "Genome-wide profiling of diel and circadian gene expression in the malaria vector Anopheles gambiae." Proceedings of the National Academy of Sciences **108**(32): 421-430.

Schroder, E. A., M. Lefta, X. Zhang, D. C. Bartos, H. Z. Feng, Y. Zhao, A. Patwardhan, J. P. Jin, K. A. Esser and B. P. Delisle (2013). "The cardiomyocyte molecular clock, regulation of Scn5a, and arrhythmia susceptibility." Am J Physiol Cell Physiol **304**(10): C954-965.

Slat, E., G. M. Freeman, Jr. and E. Herzog (2013). The Clock in the Brain: Neurons, Glia, and Networks in Daily Rhythms. Circadian Clocks. A. Kramer and M. Merrow, Springer Berlin Heidelberg. **217:** 105-123.

Storch, K. F., O. Lipan, I. Leykin, N. Viswanathan, F. C. Davis, W. H. Wong and C. J. Weitz (2002). "Extensive and divergent circadian gene expression in liver and heart." Nature

417(6884): 78-83.

Stratmann, M. and U. Schibler (2006). "Properties, entrainment, and physiological functions of mammalian peripheral oscillators." J Biol Rhythms **21**(6): 494-506.

Straume, M. (2004). "DNA microarray time series analysis: automated statistical assessment of circadian rhythms in gene expression patterning." Methods Enzymol **383**: 149-166.

Trapnell, C., L. Pachter and S. L. Salzberg (2009). "TopHat: discovering splice junctions with RNA-Seq." Bioinformatics **25**(9): 1105-1111.

Trapnell, C., B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold and L. Pachter (2010). "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation." Nat Biotechnol **28**(5): 511-515.

Vollmers, C., S. Gill, L. DiTacchio, S. R. Pulivarthy, H. D. Le and S. Panda (2009). "Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression." Proceedings of the National Academy of Sciences **106**(50): 21453-21458.

Walker, J. R. and J. B. Hogenesch (2005). RNA Profiling in Circadian Biology. Methods in Enzymology. W. Y. Michael, Academic Press. **Volume 393:** 366-376.

Wichert, S., K. Fokianos and K. Strimmer (2004). "Identifying periodically expressed transcripts in microarray time series data." Bioinformatics **20**(1): 5-20.

Wijnen, H., F. Naef and M. W. Young (2005). Molecular and Statistical Tools for Circadian Transcript Profiling. Methods in Enzymology. W. Y. Michael, Academic Press. **Volume 393:** 341-365.

Wu, T. D. and S. Nacu (2010). "Fast and SNP-tolerant detection of complex variants and splicing in short reads." Bioinformatics **26**(7): 873-881.

Xu, K., Justin R. DiAngelo, Michael E. Hughes, John B. Hogenesch and A. Sehgal (2011). "The Circadian Clock Interacts with Metabolic Physiology to Influence Reproductive Fitness." Cell Metabolism **13**(6): 639-654.

Yang, R. and Z. Su (2010). "Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation." Bioinformatics **26**(12): i168-174.

Yoo, S. H., S. Yamazaki, P. L. Lowrey, K. Shimomura, C. H. Ko, E. D. Buhr, S. M. Siepka, H. K. Hong, W. J. Oh, O. J. Yoo, M. Menaker and J. S. Takahashi (2004). "PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues." Proc Natl Acad Sci U S A **101**(15): 5339-5346.

Zhang, E. E., A. C. Liu, T. Hirota, L. J. Miraglia, G. Welch, P. Y. Pongsawakul, X. Liu, A. Atwood, J. W. Huss, 3rd, J. Janes, A. I. Su, J. B. Hogenesch and S. A. Kay (2009). "A genome-wide RNAi screen for modifiers of the circadian clock in human cells." Cell **139**(1): 199-210.

Zhang, R., N. F. Lahens, H. I. Ballance, M. E. Hughes and J. B. Hogenesch (2014). "A circadian gene expression atlas in mammals: Implications for biology and medicine." Proc Natl Acad Sci U S A.

Chapter 3: *Achilles* is a circadian clock-controlled gene that regulates immune function in *Drosophila*

Abstract

The circadian clock is a transcriptional/translational feedback loop that drives the rhythmic expression of downstream mRNAs. Termed "clock-controlled genes," these molecular outputs of the circadian clock orchestrate cellular, metabolic, and behavioral rhythms. As part of our on-going work to characterize key upstream regulators of circadian mRNA expression, we have identified a novel clock-controlled gene in *Drosophila melanogaster*, *Achilles* (*Achl*), which is rhythmic at the mRNA level in the brain and which represses expression of anti-microbial peptides in the immune system. *Achl* knock-down in neurons dramatically elevates expression of crucial immune response genes, including *IM1* (*Immune induced molecule 1*), *Mtk* (*Metchnikowin*), and *Drs* (*Drosomysin*). As a result, flies with knocked-down *Achl* expression are resistant to bacterial challenges. Meanwhile, no significant change in core clock gene expression and locomotor activity is observed, suggesting that *Achl* influences rhythmic mRNA outputs rather than directly regulating the core timekeeping mechanism. Notably, *Achl* knock-down in the absence of immune challenge significantly diminishes the fly's overall lifespan, indicating a behavioral or metabolic cost of constitutively activating this pathway. Together, our data demonstrate that (1) *Achl* is a novel clock-controlled gene that (2) regulates the immune system, and (3) participates in signaling from the brain to immunological tissues.

Introduction

Circadian rhythms are internal timekeeping mechanisms that orchestrate daily oscillations of behavior, metabolism and physiology. In most living organisms, circadian rhythms play a profound role in the regulation of physiological behaviors, such as locomotor activity, sleep-wake cycle, body temperature, blood pressure, cardiovascular activity, muscle strength, feeding,

glucose and lipid homeostasis, and alertness (Hastings et al. 2003). In addition, circadian rhythms regulate both adaptive and innate immunity and thereby influence resistance to infection (Scheiermann et al. 2013). Circadian rhythms are important for maintaining homeostasis by anticipating and adapting to predictable environmental changes. Consequently, disruption of circadian rhythms influences multiple pathologies, such as neurodegenerative diseases, cardiovascular diseases, obesity, diabetes, cancer, and depression (Hastings et al. 2003, Knutsson 2003, Klerman 2005, Halberg et al. 2006, Levi and Schibler 2007, Wulff et al. 2010).

At the molecular level, circadian rhythms are regulated by core clock genes that underlie self-sustained 24-hour feedback loops. In flies, two transcription factors, CLOCK (CLK) and CYCLE (CYC) compose the positive branch of the feedback loop while PERIOD (PER) and TIMELESS (TIM) compose the negative branch. CLK and CYC form heterodimers and bind to E-box elements located in the promoter regions of *per* and *tim*, promoting their expression. Once translated, PER and TIM dimerize and translocate into the nucleus, where they prevent CLK and CYC heterodimers from accessing E-box elements, thus decreasing the mRNA expression of *per* and *tim*. The degradation of PER and TIM resets the clock and thereby starts a new round of CLK and CYC activation. Well-studied protein modifications impose appropriate delay mechanisms, thus generating a transcriptional-translational feedback loop (TTFL) that occurs about every 24 hours (Ko and Takahashi 2006, Allada and Chung 2010, Hardin 2011).

In addition to promoting *per* and *tim* expression, CLK and CYC further drive the rhythmic expression of hundreds to thousands of downstream genes. Termed “clock-controlled genes (CCGs),” these rhythmic mRNAs are not involved in the core time-keeping mechanism but instead regulate physiological processes (Hastings et al. 2003). While the core clock genes are conserved in different tissues, CCGs are highly tissue-specific. Rhythmic transcriptome profiling

in 12 different mouse organs shows little overlap of CCGs between different tissues, as expected, given how diverse these different tissues are in their physiological demands (Zhang et al. 2014). The observation that CCGs are largely tissue-specific is seen in flies as well as mammals, suggesting that it is a well-conserved aspect of circadian output pathways (Ceriani et al. 2002, Panda et al. 2002, Storch et al. 2002, Xu et al. 2011). Consequently, the disruption of core clock genes causes systematic rhythmic disorders, while the disruption of CCGs is more likely to be linked to local disorders (Hughes et al. 2012, Jeyaraj et al. 2012). Since the outputs of the circadian clock are ultimately responsible for the clock's influence on health and physiology, it is thus necessary to identify tissue-specific CCGs and to understand their regulatory mechanisms. For example, the study of cardiac specific CCGs revealed a role of rhythmic iron channels in arrhythmia development and susceptibility (Jeyaraj et al. 2012, Schroder et al. 2013). To this end, high-throughput microarray and RNA-sequencing (RNA-seq) have greatly accelerated our understanding of CCGs in diverse tissues as well as multiple cell types (McDonald and Rosbash 2001, Keegan et al. 2007, Filichkin and Mockler 2012, Hughes et al. 2012, Menet et al. 2012, Du et al. 2014). Furthermore, the use of these high-throughput approaches in genetically modified animals enables the understanding of their regulatory mechanisms as well as the contributions of principle oscillator and peripheral oscillators to the regulation of specific CCGs (Rey et al. 2011, Xu et al. 2011, Bugge et al. 2012, Koike et al. 2012, Meireles-Filho et al. 2014, Menet et al. 2014). On-going studies in our lab and others are aimed at identifying CCGs and understanding how they mediate clock output of physiological processes related to disease and therapeutics.

In animals, circadian rhythms are regulated in hierarchy. In both mammals and insects, there are neuron-based primary oscillators located in the brain. The primary oscillator in mammals resides

in the suprachiasmatic nuclei (SCN), and in flies it is distributed among several diffuse clusters of neurons (Herzog 2007, Nitabach and Taghert 2008). In addition to the primary oscillator, there are multiple peripheral tissues that behave rhythmically. The principal oscillator integrates environmental signals and sends synchronizing cues to peripheral tissues through mechanisms that are the subject of active investigation (Hastings et al. 2003).

Circadian control of immunological defenses is one of the most dramatic examples of a pathway through which the circadian clock influences organismal health and fitness. In mammals, both principal arms of the immune system – innate and adaptive – are regulated by circadian rhythms. This is seen at both a molecular and cellular level (Silver et al. 2012a, Silver et al. 2012b, Curtis et al. 2015). High-throughput analyses have revealed rhythmicity in many genes involved in the immune response (Keller et al. 2009). In addition, cytokines and chemokines, such as IL6 (interleukin 6), TNFα (tumor necrosis factors alpha) and CXCL 12 (Chemokine (C-X-C Motif) Ligand 12) are released into the circulation in a rhythmic manner. White blood cells, including T lymphocytes, natural killer cells, macrophages, monocytes and the precursor haematopoietic stem cells are released into the circulation in a rhythmic manner and respond to stimuli rhythmically (Mendez-Ferrer et al. 2008, Lange et al. 2010, Gibbs et al. 2012, Scheiermann et al. 2013, Labrecque and Cermakian 2015, Ella et al. 2016). Together, these molecular and cellular rhythms influence organismal immunobiology in profound ways. Mice show differential resistance against infection at different times of the day. Inflammation, immune resistance, and the severity of autoimmune diseases are also found to vary throughout the day in a rhythmic manner (Cutolo 2012, Gibbs and Ray 2013, Curtis et al. 2014, Carter et al. 2016). The chronic disruption of circadian rhythms, including sleep deprivation, shift work, and jet lag can precipitate disease even in healthy individuals and exacerbate existing diseases, particularly

inflammatory conditions (Ranjbaran et al. 2007).

Drosophila has been widely used as a model organism to study the mechanisms of immune response due to its relative simplicity and its genetic tractability. Some exotic defense mechanisms notwithstanding (Watson et al. 2005), the humoral immune system in *Drosophila* is highly conserved at a molecular level with its mammalian counterpart (Muller et al. 2008). The initial discovery of an immunological role for *Toll* in *Drosophila* revolutionized the study of mammalian pattern recognition receptors (Anderson 2000, Kimbrell and Beutler 2001, Hoffmann 2003). The humoral immune system of *Drosophila* is divided into two major pathways, *Toll* pathway and *Imd* (*immune deficiency*) pathway. These two pathways combat different types of bacterial and fungal infections by distinguishing the pathogen-associated molecular patterns through pattern recognition proteins, activating the downstream AMPs (anti-microbial peptides) within the immune system, particularly within the fat body. AMPs are then secreted into the heamolymph to clear the infected pathogen (Imler and Hoffmann 2000, Hoffmann 2003). There are seven AMP families characterized in *Drosophila*: Drosomycin, Metchnikowin, Cecropins, Defensin, Attacins, Diptericin and Drosocin (Hetru et al. 2003). Similar to mammals, the immune response in *Drosophila* is also found to be rhythmic. Genes involved in immune response are rhythmically expressed, and flies infected with pathogenic bacteria at different times of the day show a rhythmic resistance peaking during the late night (McDonald and Rosbash 2001, Lee and Edery 2008, Stone et al. 2012). However, it is unclear how this is regulated at both a molecular and cellular level.

Here we show that *CG17386*, a previously uncharacterized clock-controlled gene is highly rhythmic in the fly head. Using whole-transcriptome profiling, we find that *CG17386* represses the expression of immune responsive genes. Neuron-specific *CG17386* knock-down results in

dramatically elevated levels of crucial immune response genes, including AMPs. As a result, flies with knocked-down *CG17386* expression are more resistant to immune challenge with bacteria. Notably, *CG17386* knock-down in the absence of immune challenge significantly diminishes the fly's overall lifespan, indicating an energetic or metabolic cost of constitutively activating this pathway. Hereafter we refer to *CG17386* as *Achilles* (*Achl*), in recognition that its mutant phenotype protects flies against injury and infection, while simultaneously shortening their lifespan.

Achl is a clock-controlled gene with rhythmic mRNA expression in the fly head

Previous high-throughput analyses revealed that *Achl* mRNA is rhythmically expressed in the fly head (Keegan et al., 2007) and brain (Figure 3.2A) (Hughes et al., 2012). We verified these observations using quantitative PCR (qPCR) assays on fly heads collected every four hours either in 12 hour light: 12 hour dark (LD) conditions or in constant darkness (DD). As expected, we found rhythmic Achl mRNA expression (Figure 3.1A and 3.1B) in both conditions, with damped amplitude under DD. Notably, the amplitude of *Achl* mRNA under both conditions is comparable with the amplitude of core clock genes, including *timeless*. Furthermore, ChIP-seq studies from other labs suggested that CLOCK and CYCLE directly bind to the E-box regions of the Achl promoter (Abruzzi et al., 2011; Meireles-Filho et al., 2014), and *Achl* rhythmicity is eliminated in *period* mutant flies (Figure 3.2A) (Hughes et al., 2012); its expression is also down-regulated in *Clk* mutant flies (Figure 3.2B) (McDonald and Rosbash, 2001). Taken as a whole, these results suggest that *Achl* is under the direct control of the molecular clock. *Larp7*, a homolog of *Achl* in mammals, shares a conserved RNA binding domain with *Achl* (Figure 3.3B). Since *Larp7* is rhythmically expressed in the mouse kidney (Figure 3.3A) (Zhang et al., 2014),

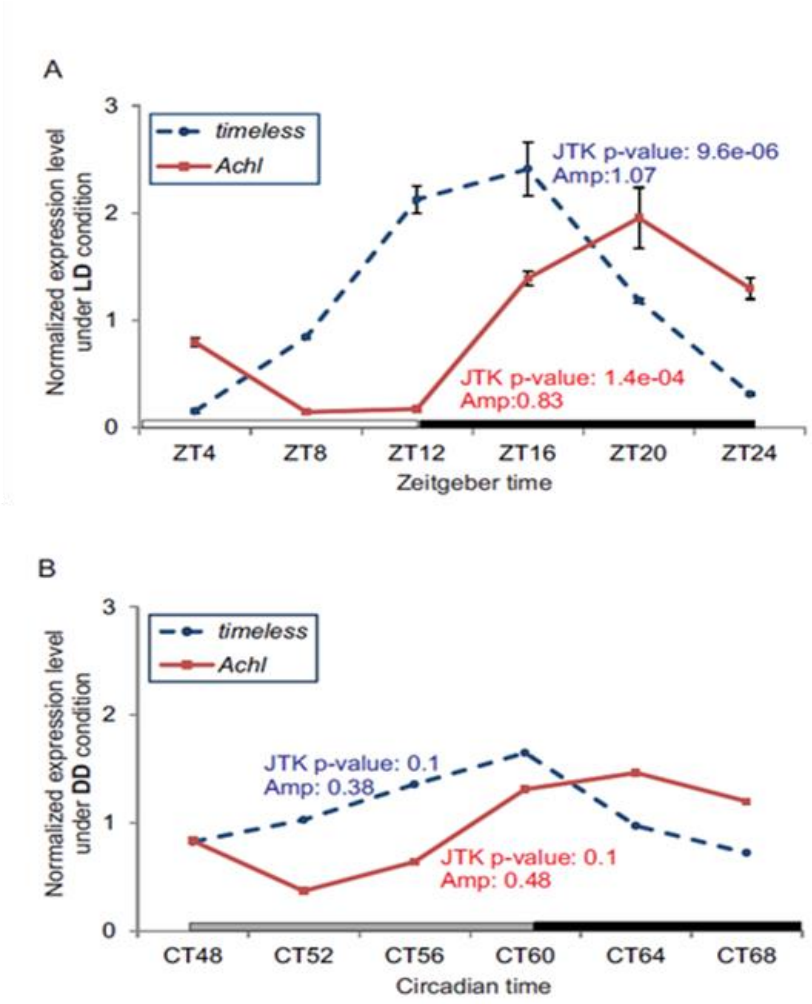
we speculate that there may be a conserved mechanism underlying their rhythmic expression. Moreover, given the role that many RNA-binding proteins play in regulating mRNA expression, these genes may in turn regulate downstream CCGs (Morf et al., 2012; Siomi and Dreyfuss, 1997).

Knocking down *AchI* in neurons does not affect the core clock

Based on the robust, high-amplitude rhythmicity of *AchI* in the fly head, we hypothesized that *AchI* is either a core clock gene or a key output gene. To test these possibilities, we generated pan-neuronal knock-down of *AchI* using genetically encoded RNAi constructs (i.e. *Elav-Gal4*; *UAS-Dcr2* was crossed with TRiP *UAS-AchI RNAi* line). qPCR data confirmed a 60% knock down efficiency in both male and female *AchI* RNAi flies at ZT2 when the mRNA level of *AchI* is moderate (Figure 3.6A), compared to the negative control, an unrelated RNAi construct with the same backbone vector and genomic insertion point (*UAS-GFP* flies).

The *AchI* RNAi line we used was generated by TRiP project and is predicted to have no off-target effects (Perkins et al., 2009). Consistent with this, we found *AchI* to be the only Drosophila mRNA with greater than 19 base pairs matching the RNAi construct we used (Figure 3.4A and 3.4B). Therefore, every potential off-target hit of this RNAi construct had at least 2 base pair mismatches and showed minimal knock-down in mRNA expression compared to *AchI*. *AchI*-RNAi flies driven under a pan-neuronal driver eclosed in normal Mendalian ratios, indicating no significant developmental effects on embryonic, larval, or pupal stages. Nevertheless, we observed partially penetrant developmental defects in adult wing development. As shown in Figure 3.5, *AchI* RNAi flies have heterogeneous defects in their wing morphology, suggesting a potential role of *AchI* in wing development and / or wing spreading behavior after eclosion.

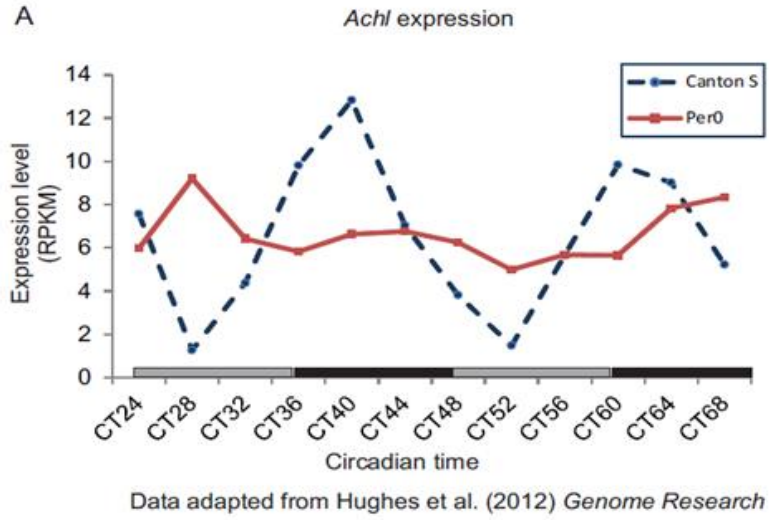
Figure 3.1. *Achl* is a clock-controlled gene that shows robust rhythmic mRNA expression in the fly head

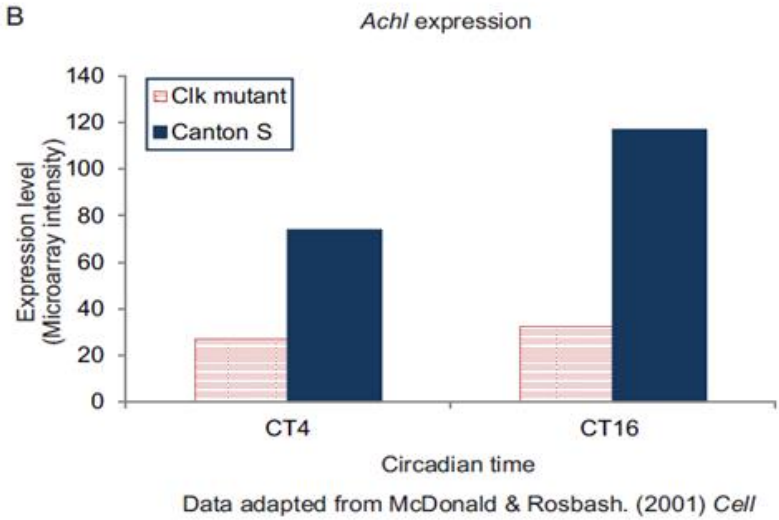


(A). qPCR assays performed on Canton S *wildtype* fly heads with two replicates showing that *Achl* is rhythmic at the mRNA level in 12hr: 12hr light:dark (LD) conditions. *timeless* is a core clock gene that serves as a positive control. The bottom horizontal white and black bars represent lights on and lights off, respectively. Error bars represent +/- SEM. Data were analyzed with JTK-CYCLE to evaluate the rhythmicity (Hughes et al., 2010; Miyazaki et al., 2011). qPCR data were normalized with the median expression value of each gene.

(B). qPCR assay performed on control fly heads showing that *Achl* is rhythmic at the mRNA level in constant darkness (DD) conditions. The bottom horizontal gray and black bars represent subjective day and night, respectively. Data were analyzed with JTK-CYCLE to evaluate the rhythmicity (Hughes et al., 2010; Miyazaki et al., 2011). qPCR data were normalized with the median expression value of each gene.

Figure 3.2. *Achl* is regulated by the core clock

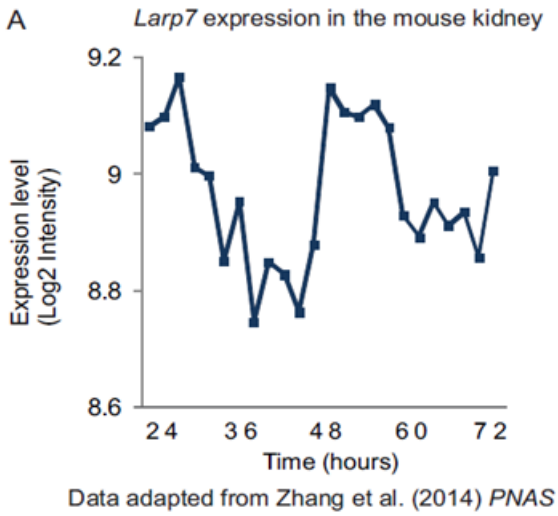




(A). *Achl* rhythmicity is disrupted in *per0* flies. Data adapted from Hughes et al. (2012) Genome Research.

(B). *Achl* expression level is down-regulated in *Clk* mutant flies. Data adapted from McDonald & Rosbash. (2001) Cell.

Figure 3.3. *Larp7*, *Achl*'s mammalian homolog is rhythmically expressed in the mouse kidney



B

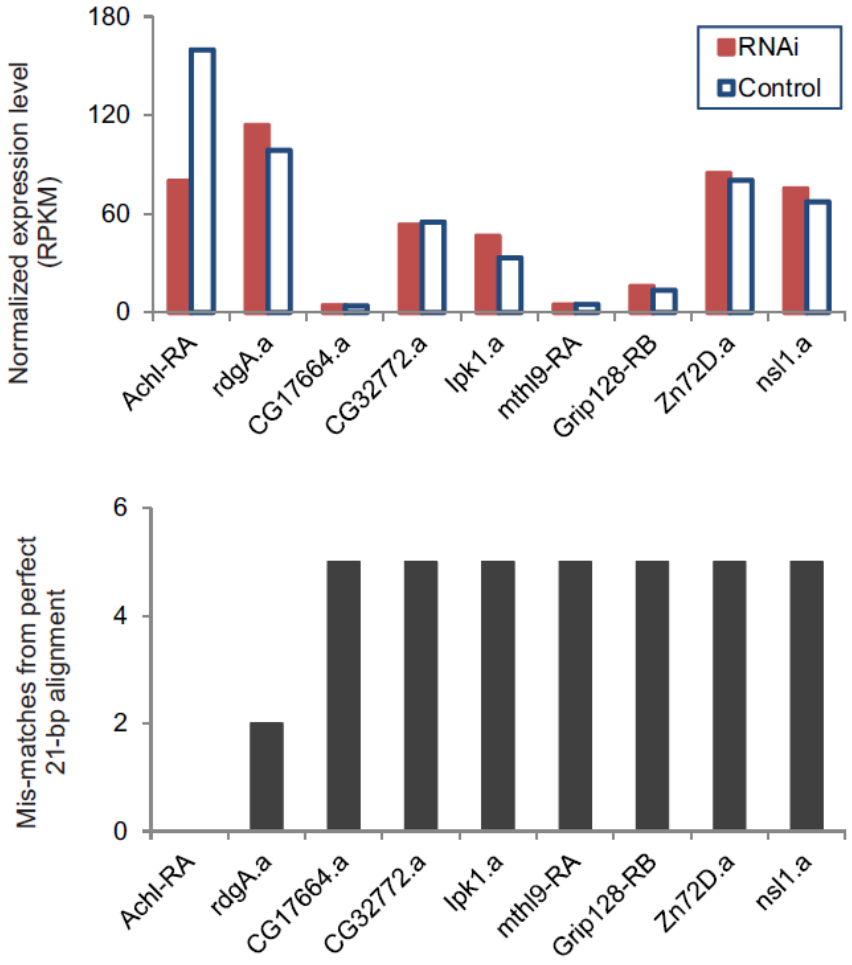
```
ACHL    EVEPPQEKEQEVPVHQEQDTEPQGPNEIDTLDPSLIPSEELAAEITDAVEFYFSNESILK
LARP7   STERKEEKKKRSRVKQ----------------------VLADIAKQVDFWFGDANLHK
        ..*  :**::   *:*                       : *:*:. *:*:*.: .: *

ACHL    DAFLLKHVRRNKEGFVSLKLVSSFKRVRQLTREWKVVGDAVRRKSRKIELNDVGTKVRRI
LARP7   DKFLREQIEKSRDGYVDISLLVSFNKMKKLTTDGKLIARAL-KSSSVVELDLEGTRIRRK
        * ** ::: ...:*:*.:.*: **:.:.:** : *::. *: ..*  :**:  **.:**

ACHL    EPLPSFDETMPSRTIVACDLPLDKLTIEKVSDLFSPCGEIALIRILKPGMAIPVDVRQFM
LARP7   KPLGERPKDEEERTVYVELLP-KNVTHSWIERVFGKCGNVVYISI--PHYKSTGDPKGFA
        :** .  :   .**: .  ** .::* ..:. :*. **::. * *  *    . * . *

ACHL    NKYPELQQKECALVEYLESSSARDARHLNGPFQVYEMVAPKKKTGKKAAVIQIAAPVARM
LARP7   FVEFETKEQAAKAIEFLNNPPEEAPRKPG--------IFPKTVKNKPIPSLQXA------
            * ::: .  :*:*:...   .*: .        : **. ..*  . :* *
```

(A). *Larp7*, a mammalian homolog of *Achl,* is rhythmically expressed in the mouse kidney. Data adapted from Zhang et al. (2014) PNAS.

(B). Sequence alignment of the conserved Lupus-La RNA binding domains in ACHL and its mammalian homolog, LARP7. The alignment is performed using MUSCLE (Multiple Sequence

Comparison by Log- Expectation) 3.8 on EMBL-EBI (Edgar, 2004). "*" indicates positions which have a single, fully conserved residue. ":" indicates conservation between groups of strongly similar properties - scoring > 0.5 in the Gonnet PAM 250 matrix, and "." indicates conservation between groups of weakly similar properties - scoring =< 0.5 in the Gonnet PAM 250 matrix. Different colors indicate the physicochemical properties of the residues.

Figure 3. 4. *Achl* RNAi knock-down is specific



(A). RNA-seq data demonstrate that the most likely off-target hits of the RNAi construct we use are not knocked-down. *Achl* is the only transcript measurably knocked-down with predicted alignment to the genetically-encoded RNAi construct we used. Control flies expressed *UAS-GFP RNAi* instead of *UAS-Achl RNAi*. RPKM: reads per kilobase per million sequenced reads.

(B). Mismatch from perfect 21-bp alignment predicted by "find OTE (off-target elements)" program (Perkins et al. 2009).

Figure 3.5. *Achl* RNAi flies have defective wing development

Pictures captured under dissection microscope.

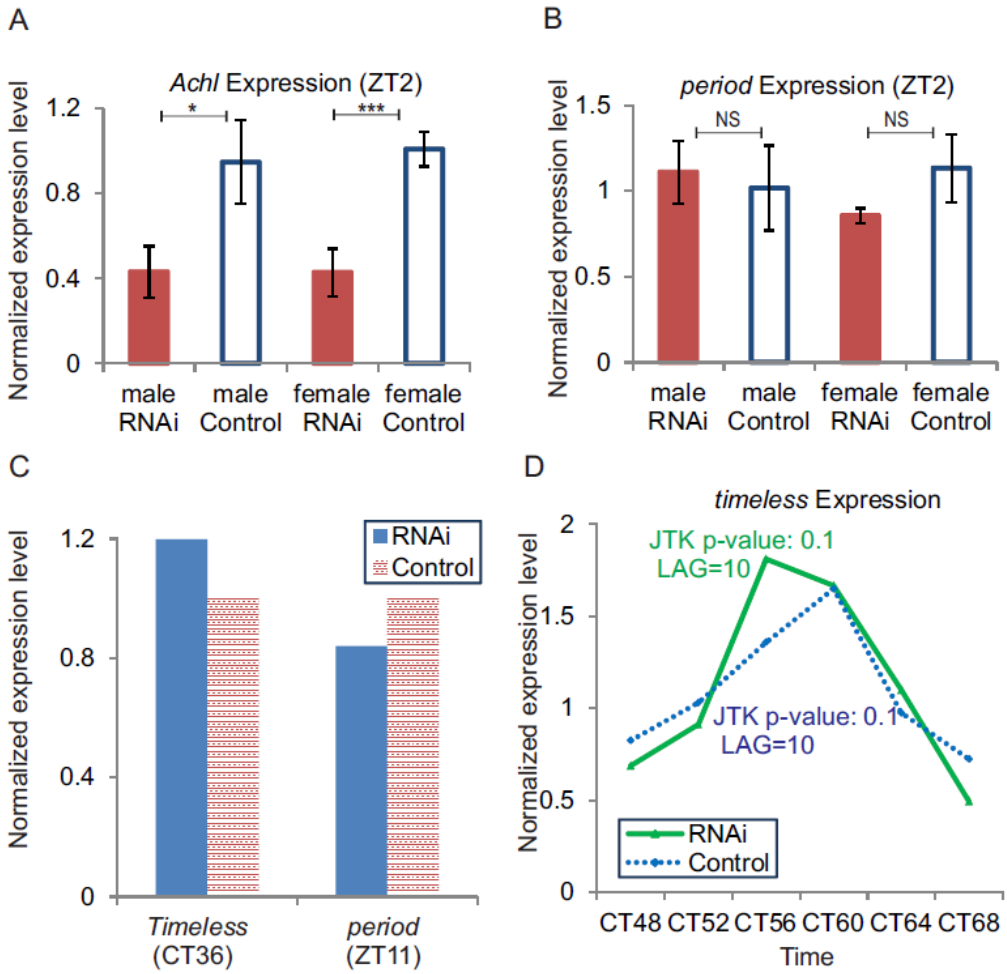
(A) Control flies (*UAS-GFP RNAi*) with normal wings.

(B) *Achl* RNAi flies with heterogeneous defects in their wing morphology, from slight to severe. Arrows indicate disrupted wing morphology.

Two hallmarks of core circadian clock genes are (1) their effect on expression of other circadian clock genes and (2) their influence on circadian period length in constant conditions. We therefore tested whether knocking down Achl has any effect on the core clock gene expression and the overall time-keeping mechanism. As shown in Figure 3.6B, there is no significant difference in the expression of a core clock gene, *per* in *Achl* RNAi flies at ZT2. Since *per* expression peaks during the night, we performed additional qPCR analyses on samples collected at ZT11 as well as CT36 with probes targeting *per* and *timeless*, two core clock genes, and found no difference in the expression level at these time points as well (Figure 3.6C), suggesting that *Achl* is unlikely to affect the core clock. Furthermore, we collected heads from flies maintained in DD every four hours for 24 hours and used qPCR to assess the rhythmicity of core clock genes. We found no obvious defects in the period, phase, or amplitude of core clock genes, as exemplified by *timeless* expression (Figure 3.6D).

In addition to the rhythmicity of core clock gene expression, we also examined the overall behavioral rhythmicity of *Achl* RNAi flies. Locomotor activity is the standard circadian behavior output that reflects the endogenous period of individual animals. We used conventional DAM (Drosophila Activity Monitoring) system monitoring to assess behavioral rhythmicity in LD and DD conditions (Pfeiffenberger et al., 2010; Schmid et al., 2011). *Achl* RNAi flies had normal locomotor rhythms in both LD and in DD (see averaged actograms in Figure 3.7A and 3.7B). The period length of these flies in constant conditions was within the experiment-to-experiment variance for negative controls (Figure 3.7A-D). The total number of rhythmic flies in the *Achl* knock-down was slightly higher than *wildtype*, although not statistically significant (Figure 3.7E). We observed that the uniformity of rhythms in *Achl* RNAi flies damped over time in DD (Figure 3.7B), and consistent with this, the power of their individual rhythms was damped on average,

Figure 3.6. Knocking down *Achl* in the neurons does not affect the core clock



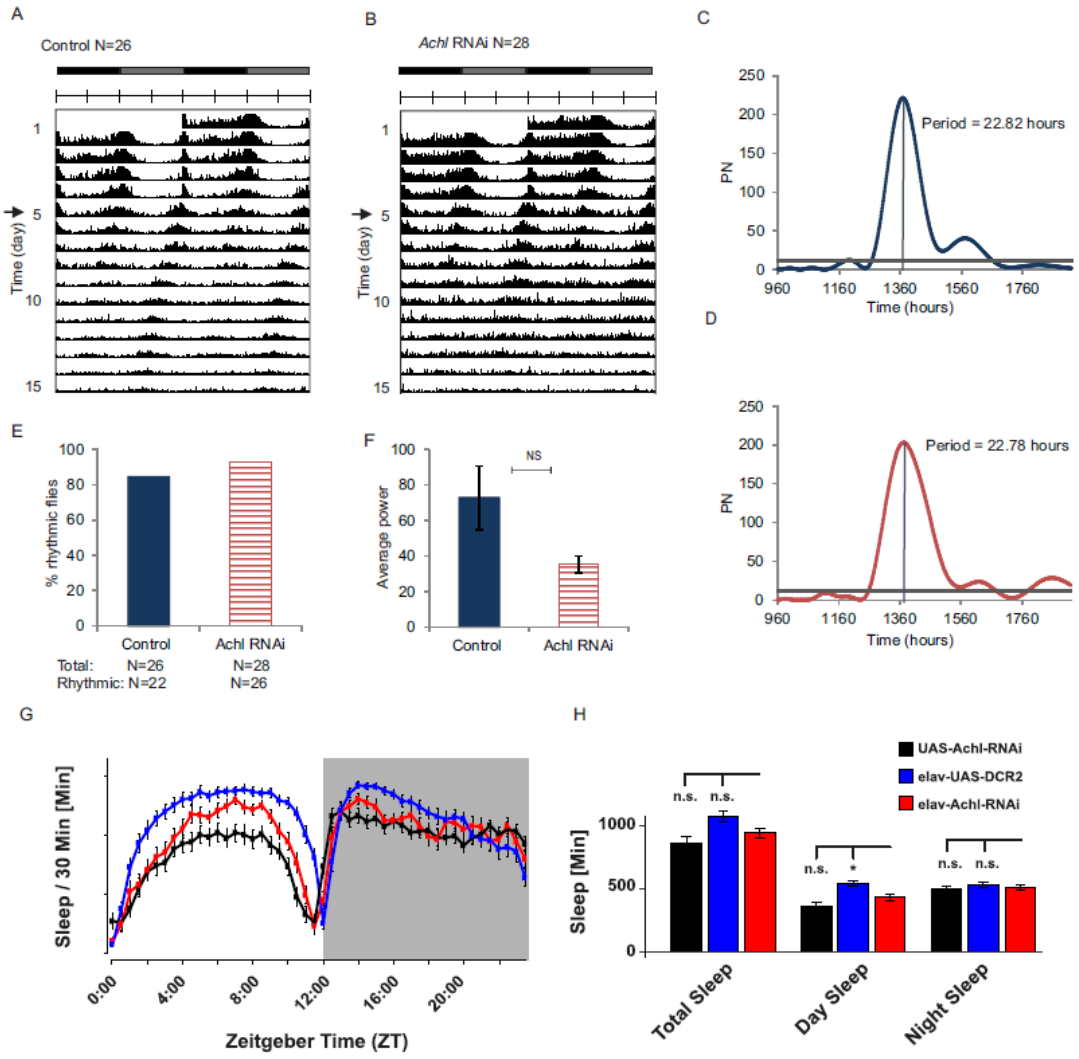
(A). qPCR data from RNAi (*UAS-Achl RNAi*) and control (*UAS-GFP RNAi*) fly heads showing that that *Achl* expression level is significantly reduced in both males and females using genetically-encoded RNAi (i.e. *Elav-Gal4; UAS-Dcr2* was crossed with TRiP *UAS-Achl/GFP RNAi* line) (* = $p < 0.05$, *** = $p < 0.005$; N = 3 biological replicates of 6-8 flies apiece; Error bars are +/- SEM). qPCR data were normalized with control *Achl* expression data of the same sex.

(B). qPCR data from RNAi and control fly heads showing that there is no significant difference in the expression of period, a core clock gene in *wildtype* control and *AchI* RNAi flies (NS = not significant, Student's T-test; N = 3 biological replicates of 6-8 flies apiece; Error bars are +/- SEM). qPCR data were normalized with control period expression data of the same sex.

(C). qPCR data from RNAi and *wildtype* control fly heads showing that there is no significant difference in the expression of core clock genes (*period* and t*imeless*) at other time points. Samples collected at CT36 were females, and samples collected at ZT11 were males. qPCR data were normalized with control expression data collected at the same time.

(D). timeless, a core clock gene, maintains its rhythmic mRNA expression in both control and RNAi fly heads, suggesting that *AchI* has no effect on the core clock. qPCR data were normalized with the median expression of each genotype.

Figure 3.7. *Achl* RNAi does not significantly affect behavioral rhythms or sleep



(A & B). Averaged actograms for representative control (N = 26) (A) and *Achl* RNAi (N = 28) (B) male flies. Flies were placed in LD conditions and converted to DD conditions at day 5, as

indicated by the arrow. These results are representative of > 5 independent experiments. The top horizontal gray and black bars represent subjective day and night.

(C & D). Lomb-Scargle periodograms corresponding to the averaged actograms in panels A and B showing the period calculated for control (C) and RNAi (D) flies. Period length differences were within the variance of negative control flies in these experiments.

(E). Percentage of individual rhythmic flies (Lomb-Scargle $p < 0.05$) observed. N = 26 for control and 28 for RNAi.

(F). Average power of individual rhythmic flies being analyzed; Error bars are +/- SEM. NS = $p > 0.05$ Students' T-test.

(G & H). Overall sleep profile (G) and sleep amounts (H) of control (parental flies) and *Achl* RNAi flies showing that *Achl* RNAi does not dramatically alter either sleep rhythms or quantity. Error bars are +/- SEM. NS = not significant, Student's T-test; *= $p < 0.05$ Students' T-test.

although not statistically significant (Figure 3.7F). The sleep profile of these flies (a key behavioral output of the circadian clock) was within the variance seen between different negative controls (Figure 3.7G-H).

Although we cannot formally exclude the possibility of a subtle behavior phenotype, based on the available data, specifically expression and rhythmicity of core clock genes, locomotor rhythms, and sleep profiles, we conclude that *Achl* is unlikely to play a direct role in the core clock. Instead, we favor the hypothesis that it is a clock-controlled gene influencing physiological outputs.

Knock-down of *Achl* results in activated expression of immune responsive genes

We then tested if *Achl* plays any role in the regulation of rhythmic physiological outputs. To this end, we profiled transcript expression using RNA-seq on fly heads collected from both male and female *Achl* RNAi and control flies to profile the genes being regulated by *Achl.* The samples were collected at ZT2, about 6 hours after the peak of *Achl* mRNA. This time point was chosen to provide sufficient time for ACHL protein to be synthesized and regulate downstream genes. We collected 3-4 biological replicates for each genotype and sex (Table 3.1). RNA samples were prepared with in-line control DNAs and quality control was performed using conventional methods. For each sample, at least five million reads were obtained, roughly consistent with previous suggestions for read-depth in *Drosophila* RNA-seq profiling (Liu et al., 2014). The raw reads were aligned against *Drosophila* genome and transcriptome with RUM (RNA-seq Unified Mapper) (Grant et al., 2011). At least 95% of sequenced reads were mapped to the genome or transcriptome. We used uniquely mapped reads to calculate expression levels and disregarded

Table 3. 1. RUM alignment statistics

Genotype	Sex	Replicate number	Total reads (millions)	Uniquely mapped reads (millions)	Total aligned reads (millions)	Transcripts with >1 RPKM	Pearson correlation coefficient transcript RPKMs (all reads)
***Achl* RNAi**	M	1	17.06	15.39 (90.2%)	16.80 (98.5%)	33208	0.9516
***Achl* RNAi**	M	2	8.34	7.24 (86.7%)	8.05 (96.5 %)	33063	
***Achl* RNAi**	M	3	5.38	4.76 (88.4%)	5.29 (98.3%)	33241	
***Achl* RNAi**	M	4	15.85	14.05 (88.7%)	15.14 (95.5%)	33414	
Control	M	1	6.02	5.47 (90.9%)	5.92 (98.4%)	33276	0.9419
Control	M	2	10.61	9.06 (85.4%)	10.38 (97.8%)	33014	
Control	M	3	9.92	8.84 (89.1%)	9.80 (98.7%)	33195	
Control	M	4	9.98	8.35 (83.6%)	9.79 (98%)	33078	
***Achl* RNAi**	F	1	5.64	5.06 (89.7%)	5.56 (98.6%)	32989	0.9719
***Achl* RNAi**	F	2	9.13	7.70 (84.3%)	8.90 (97.4%)	32851	
***Achl* RNAi**	F	3	5.64	5.00 (88.7%)	5.57 (98.6%)	32828	
Control	F	1	12.13	10.71 (88.3%)	11.90 (98%)	32704	0.9818
Control	F	2	7.61	6.77 (88.9%)	7.26 (95.3%)	32892	
Control	F	3	10.07	9.17 (91.0%)	9.96 (98.9%)	32696	

This table shows the samples we used for RNA-seq, the total number of reads, and the number of uniquely- or non-uniquely aligning reads as determined by RUM.

ambiguous reads mapping to multiple locations. The expression values were calculated with RPKM (reads per kilobase per million mapped reads). The averaged Pearson correlation coefficient for overall transcript RPKMs was greater than 0.94 for each pairwise comparison of replicates, suggesting that the data we obtained are reproducible (Table 3.1). Differentially expressed genes were analyzed with a two-way ANOVA statistical analysis with factors of genotype and sex and an explicit false-discovery correction. As shown in Table 3.2, we found dozens to hundreds of differentially expressed transcripts under different p-value thresholds that depend on genotype in both males and females. For all further analyses, we chose to use a p-value threshold of 0.001, corresponding to a false discovery rate of $q < 0.18$. As shown in Figure 3.8A, over 90% differentially-expressed transcripts are up-regulated in *Achl* RNAi flies, suggesting that ACHL's molecular function may be to repress target mRNAs.

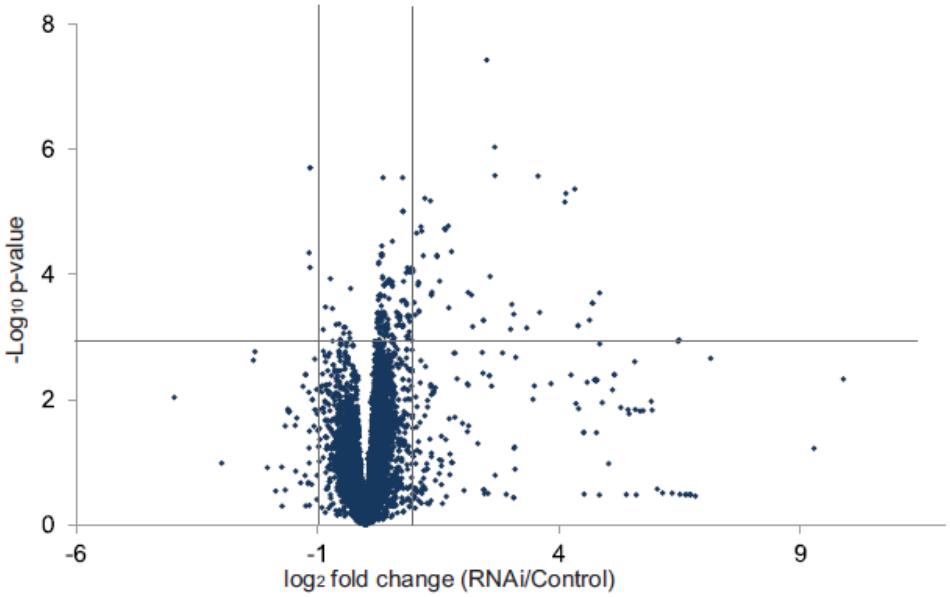
To determine the physiological pathways that are being affected, we performed Gene Ontology (GO) analysis for all the differentially expressed genes (Eden et al., 2007; Eden et al., 2009). As shown in Figure 3.8B, immune responsive pathways are dramatically enriched (see Table 3.3 for the full list). Anti-microbial peptides (AMPs) are among the well-studied immune defensive mechanisms in insects (Hetru et al., 2003; Hoffmann, 2003). We found a striking up-regulation in their expression in *Achl* RNAi flies compared to negative controls (Figure 3.8C). Moreover, we verified that immune defenses were up-regulated in the thorax and abdomen as well as in the fly head (Figure 3.9), consistent with the possibility that expression of *Achl* in the brain regulates systemic responses to infection.

Achl knock-down in neurons protects flies against bacterial infection

There are two strategies that hosts, including *Drosophila* take to defend infection: resistance and

Figure 3.8. RNA-seq data showed an activation of immune responsive genes in *Achl* RNAi flies

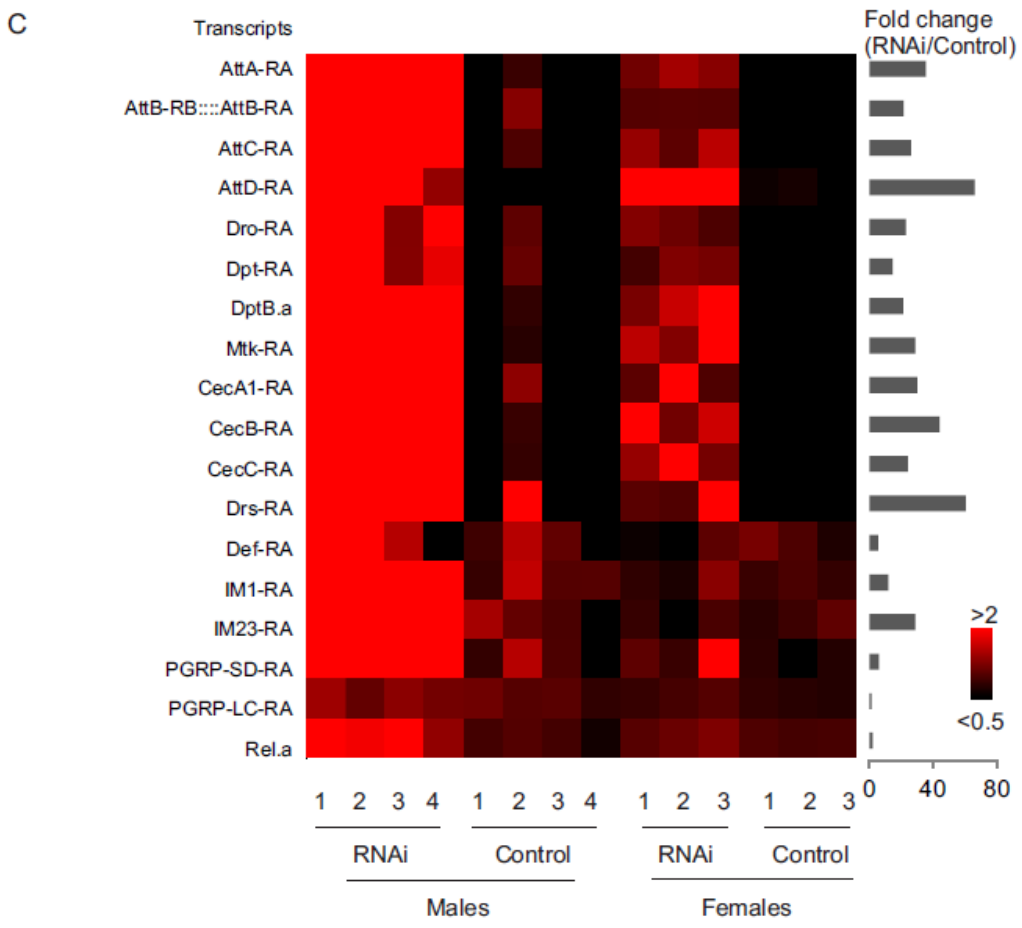
A



B

Enriched Gene Ontology terms in differentially expressed genes

Gene Ontology term	FDR q-value	Enrichment
Defense response	2.8E-4	5.86
Response to external biotic stimuli	1.73E-4	6.48
Antibacterial humoral response	1.45E-4	32.23
Response to bacterium	1.47E-4	7.94
Antimicrobial humoral response	3.6E-3	13.19
Humoral immune response	1.41E-3	11.95



(A). Volcano plot of the RNA-seq expression data. Each dot represents a single transcript. Gray dotted lines indicate a p-value threshold of < 0.001 (corresponding to a q-value < 0.18) and fold change threshold of > 2X. For each genotype, data includes four male replicates and three female replicates.

(B). Summarized Gene Ontology analysis of differentially-expressed transcripts. Software used for analysis: Gene Ontology enrichment analysis and visualization tool (GOrilla) (Eden et al. 2007, Eden et al. 2009).

(C). Heatmap of the median-normalized expression of key immune responsive genes. Red

indicates higher expression, and black indicates lower expression. Names of transcripts are marked on the left, and fold change (RNAi/Control) are marked on the right.

Table 3.2. Differential expression statistics

p-value	0.0001	0.0005	0.001	0.005	0.01	0.05
Number of transcripts	48	123	186	545	859	2490
Number of up-regulating transcripts (%)	44 (91.7%)	115 (93.5%)	171 (91.9%)	453 (83.1%)	699 (81.4%)	1775 (71.3%)
Number of genes	28	70	103	295	449	1179

This table shows the numbers of transcripts and genes differentially expressed at different p-value thresholds as well as the number and percentage of transcripts upregulated.

Table 3.3. Full Gene Ontology enrichment table

GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)
GO:0019731	antibacterial humoral response	2.35E-8	1.45E-4	32.23 (11749,27,81,6)
GO:0051707	response to other organism	7.44E-8	2.3E-4	6.55 (11749,288,81,13)
GO:0043207	response to external biotic stimulus	8.39E-8	1.73E-4	6.48 (11749,291,81,13)
GO:0009607	response to biotic stimulus	9.09E-8	1.41E-4	6.44 (11749,293,81,13)
GO:0009617	response to bacterium	1.19E-7	1.47E-4	7.94 (11749,201,81,11)
GO:0006952	defense response	2.71E-7	2.8E-4	5.86 (11749,322,81,13)
GO:0051704	multi-organism process	3.22E-7	2.84E-4	5.27 (11749,385,81,14)
GO:0006959	humoral immune response	1.82E-6	1.41E-3	11.95 (11749,85,81,7)
GO:0006950	response to stress	2.84E-6	1.95E-3	3.03 (11749,1004,81,21)
GO:0019730	antimicrobial humoral response	5.82E-6	3.6E-3	13.19 (11749,66,81,6)
GO:0098542	defense response to other organism	1.56E-5	8.79E-3	6.10 (11749,214,81,9)
GO:0009605	response to external stimulus	2.33E-5	1.2E-2	3.65 (11749,556,81,14)
GO:0042742	defense response to bacterium	2.52E-5	1.2E-2	6.67 (11749,174,81,8)
GO:0006955	immune response	4.56E-5	2.02E-2	6.14 (11749,189,81,8)
GO:0050896	response to stimulus	1.19E-4	4.89E-2	2.25 (11749,1486,81,23)
GO:0002376	immune system process	1.69E-4	6.54E-2	5.09 (11749,228,81,8)

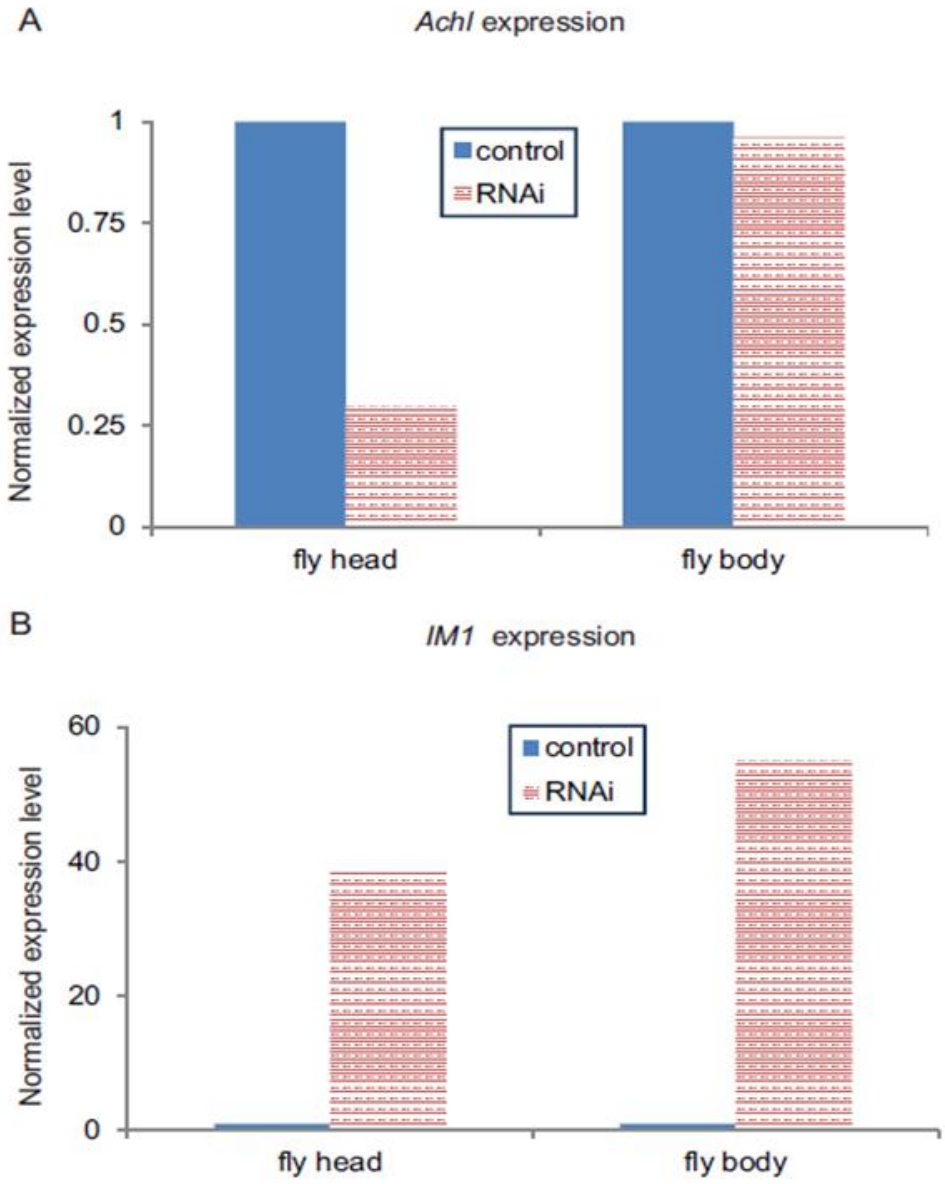
This is a full Gene Ontology enrichment table generated by GOrilla software available online

(Eden et al. 2007, Eden et al. 2009). The corresponding simplified table is shown in Figure 3B.

As shown from the GOrilla website, Enrichment (N, B, n, b) is defined as follows:

N: the total number of genes; B: the total number of genes associated with a specific GO term; n: the number of genes in the top of the user's input list or in the target set when appropriate; b: the number of genes in the intersection.

Figure 3.9. Systemic effects of *AchI* knock-down



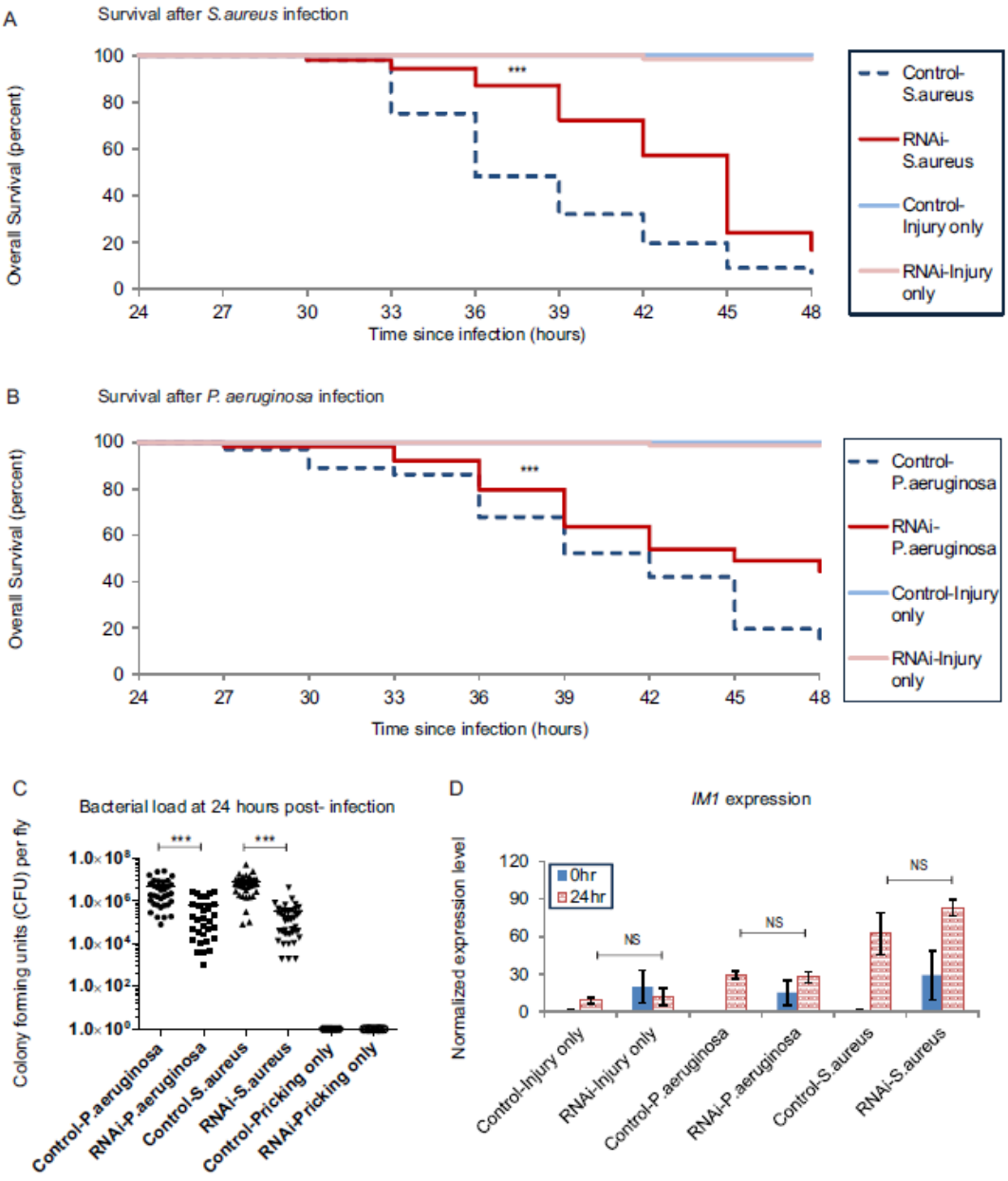
Immune response gene *IM1* is upregulated in both head and body (B) in *AchI* RNAi flies, even

though RNAi knockdown of *AchI* is only observed in head (A). qPCR data is normalized with controls of the same tissue.

tolerance. Resistance refers to the ability to reduce bacterial growth within the host body and tolerance refers to the ability to reduce detrimental pathological effects of infection. The activated expression of immune responsive genes, which play a role in reducing bacterial growth, in *Achl* RNAi flies led us to hypothesize that *Achl* RNAi flies would be more resistant toward infection. To test this, we performed a needle inoculation assay with commonly studied and evolutionarily divergent pathogenic bacteria, *P. aeruginosa* and *S. aureus*. This needle inoculation assay is a well-characterized acute infection assay that is a conventional model for testing fly immune responses (Apidianakis and Rahme, 2009). As shown in Figure 3.10A and 3.10B, *Achl* RNAi flies have better overall survival compared with control flies. In addition, we examined the growth of bacteria within the fly body, and we found decreased bacterial growth after 24 hours of infection, as indicated by fewer colony forming units (CFUs) in *Achl* RNAi flies (Figure 3.10C) with the same original bacterial load (Figure 3.11). The fact that *Achl* RNAi flies had a better overall survival and less bacterial load after 24 hours of infection compared to control flies suggests that this better survival is due to increased resistance, not tolerance.

Furthermore, we performed qPCR assays with four selected probes (*IM1, Mtk, DptB* and *PGRP-SD*) on flies infected for 24 hours to measure the expression of immune responsive genes. We chose to look at these four genes because they are genes with significant up-regulation in RNAi flies that could represent both pathways (*Toll* and *Imd* pathways) as well as different processes (bacterial recognition and anti-bacterial defense processes). As shown in Figure 3.10D & Figure 3.12, the expression of these genes after 24 hours of infection is similar for control and RNAi flies, suggesting that there is no dramatic change in the ability to respond to pathogenic bacterial infection. In addition, we found a more dramatic increased expression of these immune response genes in control flies than *Achl* RNAi flies because of the pre-activation in *Achl* RNAi flies. Taken

Figure 3.10. *Achl* knock-down in neurons protects flies against bacterial infection



(A). *Achl* knock-down flies have a higher survival rate after *S.aureus* infection. Data are the combined results of at least four biological replicates. N > 50 for infected flies and > 30 for injury

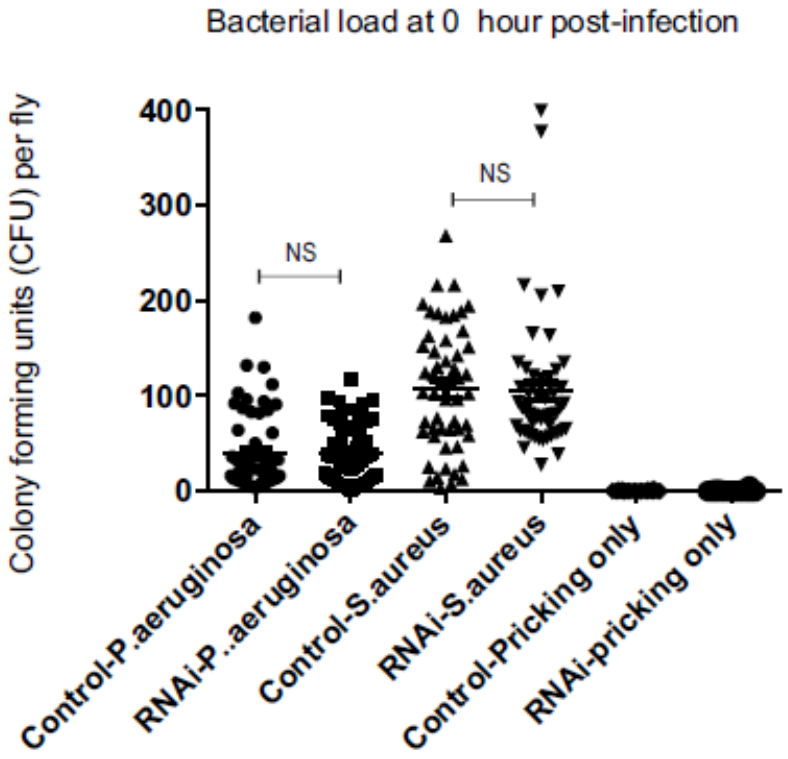
only flies. *** = p < 0.005 Log rank test.

(B). *Achl* knock-down flies have a better survival rate after *P.aeruginosa* infection. Data are the combined results of at least four biological repeats. N > 50 for infected flies and > 30 for injury only flies. *** = p < 0.005 Log rank test.

(C). *Achl* knock-down flies have decreased bacterial colony formation after bacterial infection. N > 30 for infected flies and > 20 for injury only flies. ***= p < 0.005 Students' T-test. Mean ± SEM of Control-*P.aeruginosa*: 4.898e+006 ± 1.105e+006; RNAi-*P.aeruginosa*: 606616 ± 159290; Control-*S.aureus*: 7.687e+006 ± 1.301e+006; RNAi-*S.aureus*: 324859 ± 110479.

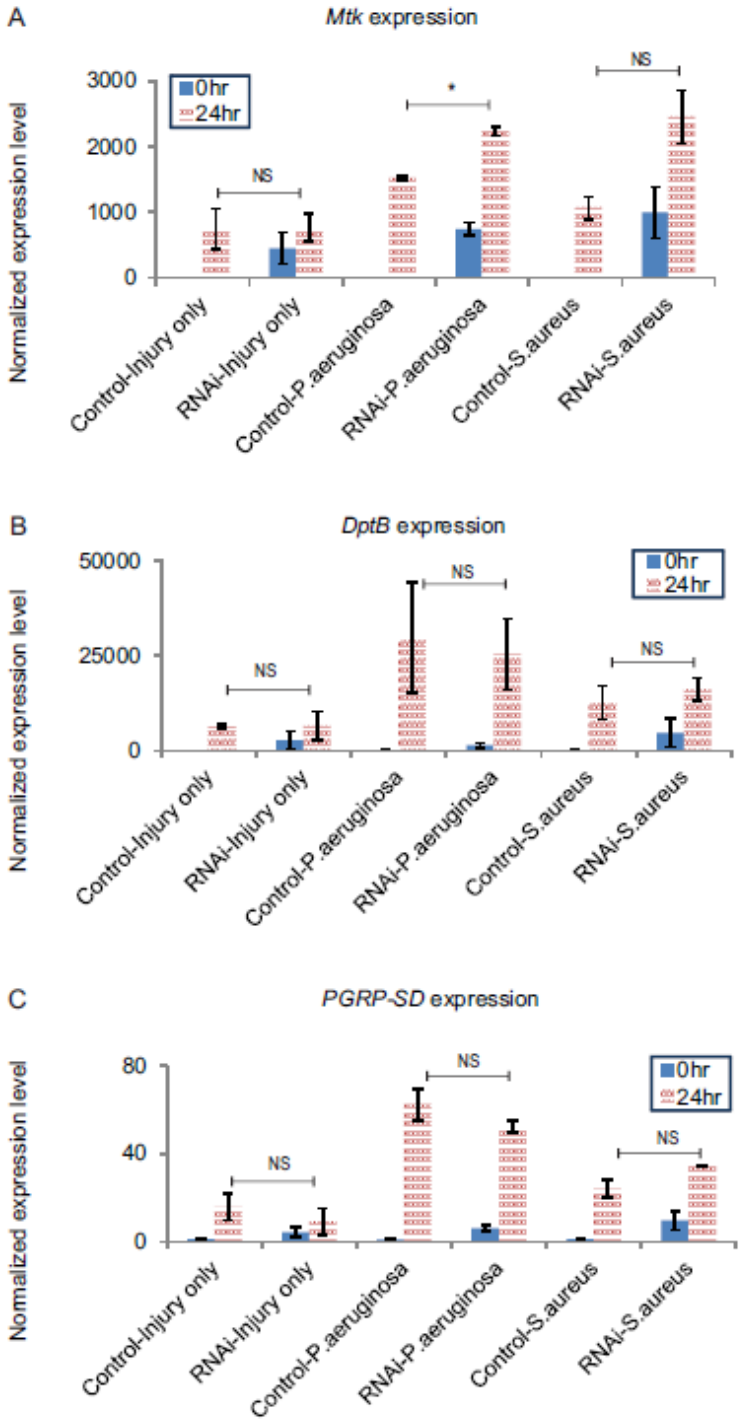
(D). Expression of *IM1* upon infection by qPCR assay (N = 2 biological replicates of 6-8 flies apiece; Error bars are +/- SEM). 0 hour expression level is higher in RNAi than control flies; 24 hours expression level is similar in RNAi and control flies for each inoculation condition. All qPCR data were normalized with control 0 hour injury only data.

Figure 3.11. Initial bacterial load of flies in the infection assay



Achl knock-down flies and control flies have comparable initial bacterial load in bacterial infection assay. N > 50 for infected flies and > 20 for injury only flies. NS = p > 0.05 Students' T-test. Mean ± SEM of Control-*P.aeruginosa*: 40.14 ± 5.408; RNAi-*P.aeruginosa*: 39.75 ± 4.216; Control-*S.aureus*: 106.8 ± 8.895; RNAi-*S.aureus*: 105.0 ± 9.471.

Figure 3.12. Expression of Immune responsive genes *Mtk*, *DptB* and *PGRP-SD* upon infection



(A-C). Expression of immune responsive genes *Mtk* (A), *DptB* (B) and *PGRP-SD* (C) upon infection examined by qPCR assay (N = 2 biological replicates of 6-8 flies apiece; Error bars are +/- SEM). 0 hour expression level is higher in RNAi than control flies; 24 hours expression level is similar in RNAi and control flies for each inoculation condition. All qPCR data were normalized with control 0 hour injury data.

together, these data suggest that *Achl* regulates steady-state immune response gene expression but does not affect peak expression after infection.

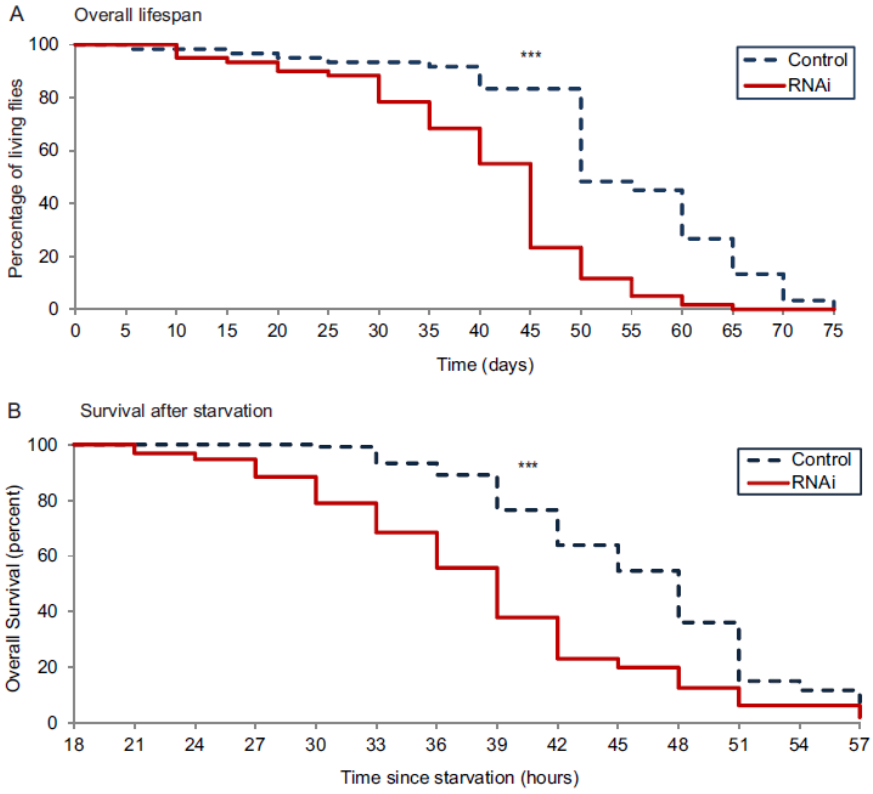
Flies with knocked-down *Achl* have a decreased lifespan

Altogether, these data indicate that *Achl* RNAi flies have a constitutively active immune system. This continuous activation is advantageous when flies are infected with bacteria, but it might cause deleterious effects in normal conditions. We therefore measured the overall lifespan of these flies and found that *Achl* RNAi flies have a significantly shorter lifespan than the control flies (Figure 3.13A). We hypothesized that this trade-off might be due to metabolic dysfunction, and therefore we examined starvation resistance of both *Achl* RNAi flies and control flies. To do so, flies were placed into fresh 1% agarose vials with water supply but without any source of nutrition. As shown in Figure 3.13B, *Achl* RNAi flies have a median survival time of 39 hours, while control flies have a median survival time of 48 hours. This decrease in both normal lifespan and starvation resistance suggests a behavioral or metabolic cost of constitutively activating immune pathways.

Discussion

The circadian clock drives tissue-specific expression of rhythmic mRNAs to regulate many different physiological processes, including the sensitivity and activity of the immune system. Here we show that *Achl* is expressed with 24-hour periodicity in the fly brain. Given the phase of its expression, the presence of tandem E-boxes with CLK/CYC occupancy in ChIP-seq studies, and disrupted rhythmic expression in Clock and period mutant flies, it is likely to be under the direct control of the molecular circadian clock. Since *Achl* encodes a protein with an RNA-binding domain, we hypothesized that it may in turn regulate the expression of downstream

Figure 3.13. *Achl* RNAi flies have a shorter lifespan and decreased starvation resistance



(A). Overall lifespan of *Achl* RNAi and control flies. Data are combined from at least three independent replicates. Total N = 60; *** = $p < 0.005$ Log rank test.

(B). Starvation assay performed with *Achl* RNAi and control flies. Data are obtained by merging at least three replicates. Total N = 90-120; *** = $p < 0.005$ Log rank test.

rhythmic genes. To test this possibility, we used whole-transcriptome RNA-seq to profile gene expression differences in the heads of *wildtype* flies and those with knocked-down *AchI* expression. Surprisingly, we found a dramatic increase in the expression of immune-response genes (Gene Ontology analysis $q < 10\text{-}4$; enrichment score > 5.0). Notably, most of these genes were up-regulated (*Table 3.2* and Figure 3.8), with expression levels ranging from 2- to 1000-fold greater than baseline. Taken together, these data suggest that AchI directly or indirectly regulates gene expression of immune effectors.

As expected from these results, knock-down of *AchI* potentiates the immune system and thereby increases flies' resistance to bacterial infection. We acknowledge that this phenotype may be due in part to changes in bacterial tolerance, although we have no data to test this possibility directly. Related to this, we also acknowledge the possibility that *AchI* knock-down may disrupt the microbiota in flies in a complicated and unpredictable fashion and might thereby alter normal immune function, for example, by increasing the basal expression of anti-microbial peptides. Similarly, phagocytosis is part of the rhythmic immune response in *Drosophila*, and its disruption might contribute to the survival phenotype we observe. Since *AchI* is not measurably expressed in hemocytes (Cherbas et al., 2011), any effect on phagocytosis is likely to be indirect and via a heretofore uncharacterized signaling pathway. Nevertheless, we cannot formally exclude the possibility that *AchI* plays a role in phagocytosis. Moreover, given the relatively minor developmental phenotype in *AchI* knock-downs compared to the enormous up-regulation of anti-microbial peptides, we favor the conclusion that the survival benefit of *AchI* knock-down is due to activation of immune response genes rather than a more elaborate mechanism. Finally, we note that *AchI* RNAi flies have shorter overall lifespans in the absence of infection, particularly in starvation conditions, thus indicating a metabolic or physiological cost of having a perpetually

activated immune system.

Achl' closest mammalian homologues are *La ribonucleoprotein domain family members 6 and 7* (*Larp6* and *Larp7*). Neither gene has a known function in immunity and neither has polymorphisms that predispose human patients to auto-immune disorders in GWAS studies (Farh et al., 2015). We note that *Larp7* cycles in the mouse kidney, and *Larp6* cycles (albeit more weakly) in the distal colon, so both genes possess promoter and/or enhancer elements necessary for clock-driven rhythmicity. However, there is a relative paucity of circadian transcriptional profiling in immunological tissues, so it is difficult to say whether either gene is rhythmic in cells directly relevant to immunity.

Several follow-up questions emerge from these observations that are the subject of on-going work in our laboratory. First, how does *Achl* regulate immune responses at a cellular level? We note that *Achl* expression is not normally detected in hemocytes or fat body, the conventional workhorses of the fly's immune system. Moreover, we note that knock-down of *Achl* in the head affects immune gene expression in the body, indicating system-wide influences of *Achl* on immune function (Figure 3.9). We further emphasize that our experiments manipulated *Achl* expression strictly in neurons. Since expression of anti-microbial peptides is mainly a product of the fly fat body, we surmise that there must be a signaling mechanism downstream of *Achl* that conveys information from neurons to the fat body or other immunologically relevant tissues. Given the effect of *Achl* on the fly's response to starvation, we speculate that signaling downstream of *Achl* influences both immune function and metabolic control, although more complicated mechanisms are certainly possible. Testing these possibilities and linking *Achl* to circadian control of both energy expenditure and immune function is a priority of on-going studies.

Second, does *Achl* mRNA cycling directly contribute to functional rhythmicity in anti-bacterial immune defenses? We note that roughly 100-200 genes in the fly nervous system are under circadian control, and that many of them, like *Achl*, have never been studied in detail. Given the importance of understanding the molecular mechanisms by which circadian clocks control immune function (Edgar et al., 2016; Fortier et al., 2011; Gibbs et al., 2012; Keller et al., 2009; Rahman et al., 2015; Scheiermann et al., 2013; Silver et al., 2012b), we propose the testable hypothesis that Achl acts as a direct link between the circadian clock and immune function. This hypothesis is supported by several observations: *Achl* has high-amplitude mRNA rhythms that are likely driven by the clock, and it encodes an RNA-binding protein that may in turn regulate expression of downstream circadian effectors, perhaps including upstream regulators of immune activation. Arguing against this hypothesis are the observations that many clock gene mutants disrupt rather than enhance anti-bacterial immunity (Lee and Edery, 2008; Stone et al., 2012). However, we observe that *Clk* mutants actually confer enhanced resistance to infection (Lee and Edery, 2008), and we further note that *Clk* mutants have diminished *Achl* expression (McDonald and Rosbash, 2001). Based on this molecular phenotype, our data would predict increased expression of AMPs in *Clk* mutants, due to their reduced expression of *Achl.* This prediction is confirmed by previous microarray studies (McDonald and Rosbash, 2001). We acknowledge that there are a multitude of potential molecular mechanisms that may account for *Achl*'s somewhat anomalous immunological phenotype. But, given the enormous contributions flies have made as a model system to both immunity and circadian rhythms, we contend that testing this hypothesis will contribute to understanding the fundamental mechanisms linking circadian rhythms and immune function in all animals.

Materials and Methods

Fly stocks and behavioral monitoring

Flies were maintained on standard food (Genesee Scientific, San Diego, California) at 25 °C in 12 hour light: 12 hour dark (LD) conditions. Humidity was maintained at roughly 50%. All fly stocks used were acquired from the Bloomington stock center: *Elav-Gal4* strain: *P{w[+mW.hs]=GawB}elav[C155] w[1118]; P{w[+mC]=UAS-Dcr-2.D}2*. *CG17386* (*Achl*) RNAi strain: *y[1] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.JF01976}attP2*. RNAi control strain: *y[1] v[1]; P{y[+t7.7] v[+t1.8]=UAS-GFP.VALIUM10}attP2*. Canton S flies were used as *wildtype* controls.

Individual male flies were placed in locomotor activity monitor tubes 3-5 days after eclosion and were entrained to LD conditions for five days before being released into free-running conditions of constant darkness (DD). Automated TriKinetics (Waltham, Massachusetts) infrared beam-crossing monitor systems were used to assay locomotor activity using one minute bins.

Bacterial stocks and culture

P. aeruginosa strain PAO1 was a gift from Dr. Lon Chubiz (UMSL). *S. aureus* strain was acquired by Kelly O'Mara from Carolina Biological (Burlington, North Carolina). For each experiment, a frozen glycerol stock was freshly streaked onto a LB plate and grown overnight at 37 °C. A single colony was picked from this plate and grown in 1-2 ml LB media overnight. Afterwards a subculture was made in 2 ml LB at a starting OD600 nm of 0.05 or less. The culture was harvested at an OD600 nm of around 3.0. After 1X PBS wash, the bacterial was serially diluted into OD600 nm of 0.05 for *P. aeruginosa* (about 40 bacteria/fly) or 0.10 for *S. aureus* (about 100 bacteria/fly) with 1X PBS for experimental infection.

RNA Preparation

Flies of three to five days old were used for sequencing and expression analysis of *Achl* core clock genes. Five to ten days old flies were used for infection related qPCR analysis. For sequencing and quantitative PCR purposes, five to ten fly heads per sample were manually dissected in PBS, transferred into 100 µL of Trizol (Life Technologies, Carlsbad, California) and homogenized using RNase-free pestles (Fisher Scientific, Waltham, Massachusetts). After 5 min incubation at room temperature with an additional 400 µL of Trizol, total RNA was prepared with phase lock gels (5Prime, Gaithersburg, Maryland) and RNeasy mini kit (Qiagen, Hilden, Germany) using the manufacturer's protocol. RNA quantity and quality were assessed using a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, California) and a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California).

Library preparation and RNA sequencing

RNA-seq libraries were prepared from 200ng total RNA per sample using TruSeq Stranded mRNA LT Sample Prep Kit (Illumina, San Diego, California) following the manufacturer's protocol with 13 rounds of PCR amplification. Libraries were quantified and qualified using Qubit 2.0 Fluorometer and Agilent 2100 Bioanalyzer. Prepared libraries with unique indexes were diluted to 2.5nM and multiplexed for loading on an Illumina Miseq (UMSL Genomics Facility) for sequencing using Miseq reagent kit v2 (50 cycles). Sequencing samples with quality scores >= Q30 were over 95%.

RNA-seq alignment

The RNA-seq Unified Mapper (Grant et al., 2011) was used to align sequenced reads to the genome and transcriptome of Drosophila melanogaster (build dm3) using the following parameters:

"--strand-specific --variable-length-reads --bowtie-nu-limit 10 --nu-limit 10". All aligned samples showed expected proportions of uniquely and non-uniquely aligned reads, and normal distribution across Drosophila chromosomes, as compared to previous studies (Hughes et al., 2012). Pearson correlation coefficient of all transcripts is no less than 0.94 between replicates.

Quantitative and reverse transcriptase PCR

Reverse transcriptase PCR was performed using High Capacity cDNA Reverse Transcription kit (Applied Biosystems, Foster City, California) following the manufacturer's protocol with 100 nanograms to 500 nanograms of RNA. Quantitative PCR (qPCR) was performed using Brilliant III ultra-fast qPCR master mix kit (Agilent Technologies, Santa Clara, California) on an MX3005p qPCR system (Agilent Technologies, Santa Clara, California) with 10–50 ng of cDNA template. All the probes were Taqman probes labeled with FAM (fluorescein) (Applied Biosystems, Foster City, California). The Taqman probes used include: *CG17386* (*Achl*), Dm01824077_s1; *period*, Dm01843684_g1;*timeless*, Dm_01814247_g1; *IM1*, Dm02366433_s1; *Mtk*, Dm01821460_s1; *DptB*, Dm01821557_g1; *PGRP-SD*, Dm01840723_s1; and *Rpl32*: Dm_02151827_g1.

Infection and survival assay

Survival rate assay protocol was performed as described previously (Apidianakis and Rahme, 2009). Briefly, male flies 1 to 2 days after eclosion were collected, raised in the incubator on standard food (Genesee Scientific, San Diego, CA) at 25 °C in LD conditions until 5 to 6 days after eclosion. At ZT2 (zeitgeber time 2, i.e. 2 hours after lights-on), an ethanol sterilized tungsten needle that is of about 0.01 mm diameter at the tip and 0.25 mm across the needle body was injected into the thorax of five to ten flies to be coated with fly hemolymph. The needle was then dipped into the diluted bacterial solution prepared earlier and injected into the midline

of the thorax of flies, as described above. Control flies were injected with 1X PBS instead of bacterial solution. Injured flies were moved into vials containing fresh food, transferred daily into new vials at 25 °C. Viability was checked every three hours between 24 hours and 48 hours post injection. For each time point and genotype, data from at least three independent experiments were pooled.

Colony forming unit assay

Colony forming unit assay protocol was adapted from the one described previously (Apidianakis and Rahme, 2009). The infection step is performed the same as in the survival rate assay described above. After 0 or 24 hours, individual flies were rinsed in 70% ethanol; homogenized in 100 ml of 1X PBS; serially diluted and spread on LB plates. These plates were incubated at 37 °C overnight before the total number of colonies was counted. For each time point and genotype, data from at least three independent experiments were pooled.

Lifespan assay

Male flies newly enclosed were collected and maintained at 25 °C in LD conditions in fresh vials with a density of 20 flies per vial. Flies were monitored for viability and transferred into new vials every five days for eight weeks. Data from three independent experiments were pooled.

Starvation assay

Seven to nine days old male flies maintained at 25 °C in LD conditions were transferred to new vials containing 1% agarose with a density of about 20 flies per vial with the same maintenance conditions. Flies were monitored for viability every three hours afterwards.

Statistical Analyses

Analysis of rhythmicity: JTK_CYCLE is an algorithm implemented in R that analyzes the rhythmicity of time-series data (Hughes et al., 2010). JTK_Cycle was performed using a period length window precisely equal to 24 hours. Lag indicates the phase of serial expression, and amplitude was calculated as previously described (Miyazaki et al., 2011).

Survival assay: Survival data were analyzed using log rank test as previously described (H. J. Motulsky, 2016).

CFU assay: The comparison of CFUs in control and RNA flies were performed using students' T-test.

qPCR: Relative expression from qPCR data were calculated using the delta delta Ct method. Briefly, Ct data were normalized with internal control (*Rpl32*: Dm_02151827_g1), then normalized with either control (for gene analysis) or median (for circadian analysis) as mentioned in the Results and Legends.

RNA-Seq: GO term analysis was performed using GOrilla software available online (Eden et al. 2007, Eden et al. 2009). Two-way ANOVA was used to compare RNAi and control flies. Transcripts with averaged RPKM<1 were eliminated from further analysis. An explicit false-discovery correction (i.e. the q-value) was calculated using the method described by Benjamani-Hochberg ((Benjamini and Hochberg 1995).

Fly Behavior: Double-plotted actograms and Lomb-Scargle periodograms for assaying free-running period were generated using ActogramJ implemented in ImageJ (Schmid et al. 2011). DD statistics were measured from the first through the tenth subjective day (DD1-DD10; i.e. days 5-15 of the experiment). Only flies surviving the length of the experiment were used for statistical analysis. Individual flies were deemed to be rhythmic if their Lomb-Scargle p-value was less

than 0.05 (Figure 3.7E). Rhythmic power was calculated by Lomb-Scargle for the averaged actograms (Figure 3.7A-D) as well as for each individual fly (Figure 3.7F). Sleep analysis was performed as described in detail previously (Kunst et al. 2014). To measure sleep, which is considered as 5 minutes of inactivity, we used a custom written MATLAB script (Parisky et al. 2008). Sleep parameters were statistically analyzed and plotted using custom written R scripts (available on request).

Data access

All raw sequencing data have been submitted to the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE80738.

Bibliography

Allada, R. and B. Y. Chung (2010). "Circadian organization of behavior and physiology in Drosophila." Annu Rev Physiol **72**: 605-624.

Anderson, K. V. (2000). "Toll signaling pathways in the innate immune response." Curr Opin Immunol **12**(1): 13-19.

Benjamini, Y. and Y. Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." Journal of the Royal Statistical Society. Series B (Methodological) **57**(1): 289-300.

Bugge, A., D. Feng, L. J. Everett, E. R. Briggs, S. E. Mullican, F. Wang, J. Jager and M. A. Lazar (2012). "Rev-erbα and Rev-erbβ coordinately protect the circadian clock and normal metabolic function." Genes & Development **26**(7): 657-667.

Carter, S. J., H. J. Durrington, J. E. Gibbs, J. Blaikley, A. S. Loudon, D. W. Ray and I. Sabroe (2016). "A matter of time: study of circadian clocks and their role in inflammation." J Leukoc Biol **99**(4): 549-560.

Ceriani, M. F., J. B. Hogenesch, M. Yanovsky, S. Panda, M. Straume and S. A. Kay (2002). "Genome-wide expression analysis in Drosophila reveals genes controlling circadian behavior." J Neurosci **22**(21): 9305-9319.

Curtis, A. M., M. M. Bellet, P. Sassone-Corsi and L. A. O'Neill (2014). "Circadian clock proteins and immunity." Immunity **40**(2): 178-186.

Curtis, A. M., C. T. Fagundes, G. Yang, E. M. Palsson-McDermott, P. Wochal, A. F. McGettrick, N. H. Foley, J. O. Early, L. Chen, H. Zhang, C. Xue, S. S. Geiger, K. Hokamp, M. P. Reilly, A. N. Coogan, E. Vigorito, G. A. FitzGerald and L. A. J. O'Neill (2015). "Circadian control of innate immunity in macrophages by miR-155 targeting Bmal1." Proceedings of the National Academy of Sciences **112**(23): 7231-7236.

Cutolo, M. (2012). "Chronobiology and the treatment of rheumatoid arthritis." Curr Opin Rheumatol **24**(3): 312-318.

Du, N. H., A. B. Arpat, M. De Matos and D. Gatfield (2014). "MicroRNAs shape circadian hepatic gene expression on a transcriptome-wide scale." Elife **3**: e02510.

Eden, E., D. Lipson, S. Yogev and Z. Yakhini (2007). "Discovering motifs in ranked lists of DNA sequences." PLoS Comput Biol **3**(3): e39.

Eden, E., R. Navon, I. Steinfeld, D. Lipson and Z. Yakhini (2009). "GOrilla: a tool for discovery

and visualization of enriched GO terms in ranked gene lists." BMC Bioinformatics **10**(1): 1-7.
Ella, K., R. Csepanyi-Komi and K. Kaldi (2016). "Circadian regulation of human peripheral neutrophils." Brain Behav Immun.
Filichkin, S. and T. Mockler (2012). "Unproductive alternative splicing and nonsense mRNAs: A widespread phenomenon among plant circadian clock genes." Biology Direct **7**(1): 1-15.
Gibbs, J. E., J. Blaikley, S. Beesley, L. Matthews, K. D. Simpson, S. H. Boyce, S. N. Farrow, K. J. Else, D. Singh, D. W. Ray and A. S. I. Loudon (2012). "The nuclear receptor REV-ERBα mediates circadian regulation of innate immunity through selective regulation of inflammatory cytokines." Proceedings of the National Academy of Sciences **109**(2): 582-587.
Gibbs, J. E. and D. W. Ray (2013). "The role of the circadian clock in rheumatoid arthritis." Arthritis Res Ther **15**(1): 205.
Halberg, F., G. Cornelissen, W. Ulmer, M. Blank, W. Hrushesky, P. Wood, R. K. Singh and Z. Wang (2006). "Cancer chronomics III. Chronomics for cancer, aging, melatonin and experimental therapeutics researchers." J Exp Ther Oncol **6**(1): 73-84.
Hardin, P. E. (2011). "Molecular genetic analysis of circadian timekeeping in Drosophila." Advances in genetics **74**: 141-173.
Hastings, M. H., A. B. Reddy and E. S. Maywood (2003). "A clockwork web: circadian timing in brain and periphery, in health and disease." Nat Rev Neurosci **4**(8): 649-661.
Herzog, E. D. (2007). "Neurons and networks in daily rhythms." Nat Rev Neurosci **8**(10): 790-802.
Hetru, C., L. Troxler and J. A. Hoffmann (2003). "Drosophila melanogaster antimicrobial defense." J Infect Dis **187 Suppl 2**: S327-334.
Hoffmann, J. A. (2003). "The immune response of Drosophila." Nature **426**(6962): 33-38.
Hughes, M. E., G. R. Grant, C. Paquin, J. Qian and M. N. Nitabach (2012). "Deep sequencing the circadian and diurnal transcriptome of Drosophila brain." Genome Res **22**(7): 1266-1281.
Imler, J. L. and J. A. Hoffmann (2000). "Signaling mechanisms in the antimicrobial host defense of Drosophila." Curr Opin Microbiol **3**(1): 16-22.
Jeyaraj, D., S. M. Haldar, X. Wan, M. D. McCauley, J. A. Ripperger, K. Hu, Y. Lu, B. L. Eapen, N. Sharma, E. Ficker, M. J. Cutler, J. Gulick, A. Sanbe, J. Robbins, S. Demolombe, R. V. Kondratov, S. A. Shea, U. Albrecht, X. H. Wehrens, D. S. Rosenbaum and M. K. Jain (2012). "Circadian rhythms govern cardiac repolarization and arrhythmogenesis." Nature **483**(7387): 96-99.

Keegan, K. P., S. Pradhan, J. P. Wang and R. Allada (2007). "Meta-analysis of Drosophila circadian microarray studies identifies a novel set of rhythmically expressed genes." PLoS Comput Biol **3**(11): e208.

Keller, M., J. Mazuch, U. Abraham, G. D. Eom, E. D. Herzog, H. D. Volk, A. Kramer and B. Maier (2009). "A circadian clock in macrophages controls inflammatory immune responses." Proc Natl Acad Sci U S A **106**(50): 21407-21412.

Kimbrell, D. A. and B. Beutler (2001). "The evolution and genetics of innate immunity." Nat Rev Genet **2**(4): 256-267.

Klerman, E. B. (2005). "Clinical aspects of human circadian rhythms." J Biol Rhythms **20**(4): 375-386.

Knutsson, A. (2003). "Health disorders of shift workers." Occupational Medicine **53**(2): 103-108.

Ko, C. H. and J. S. Takahashi (2006). "Molecular components of the mammalian circadian clock." Hum Mol Genet **15 Spec No 2**: R271-277.

Koike, N., S.-H. Yoo, H.-C. Huang, V. Kumar, C. Lee, T.-K. Kim and J. S. Takahashi (2012). "Transcriptional Architecture and Chromatin Landscape of the Core Circadian Clock in Mammals." Science **338**(6105): 349-354.

Kunst, M., Michael E. Hughes, D. Raccuglia, M. Felix, M. Li, G. Barnett, J. Duah and Michael N. Nitabach (2014). "Calcitonin Gene-Related Peptide Neurons Mediate Sleep-Specific Circadian Output in Drosophila." Current Biology **24**(22): 2652-2664.

Labrecque, N. and N. Cermakian (2015). "Circadian Clocks in the Immune System." Journal of Biological Rhythms **30**(4): 277-290.

Lange, T., S. Dimitrov and J. Born (2010). "Effects of sleep and circadian rhythm on the human immune system." Ann N Y Acad Sci **1193**: 48-59.

Lee, J. E. and I. Edery (2008). "Circadian regulation in the ability of Drosophila to combat pathogenic infections." Curr Biol **18**(3): 195-199.

Levi, F. and U. Schibler (2007). "Circadian rhythms: mechanisms and therapeutic implications." Annu Rev Pharmacol Toxicol **47**: 593-628.

McDonald, M. J. and M. Rosbash (2001). "Microarray analysis and organization of circadian gene expression in Drosophila." Cell **107**(5): 567-578.

Meireles-Filho, A. C., A. F. Bardet, J. O. Yanez-Cuna, G. Stampfel and A. Stark (2014). "cis-regulatory requirements for tissue-specific programs of the circadian clock." Curr Biol **24**(1): 1-10.

Mendez-Ferrer, S., D. Lucas, M. Battista and P. S. Frenette (2008). "Haematopoietic stem cell release is regulated by circadian oscillations." Nature **452**(7186): 442-447.
Menet, J. S., S. Pescatore and M. Rosbash (2014). "CLOCK:BMAL1 is a pioneer-like transcription factor." Genes Dev **28**(1): 8-13.
Menet, J. S., J. Rodriguez, K. C. Abruzzi and M. Rosbash (2012). "Nascent-Seq reveals novel features of mouse circadian transcriptional regulation." Elife **1**: e00011.
Muller, U., P. Vogel, G. Alber and G. A. Schaub (2008). "The innate immune system of mammals and insects." Contrib Microbiol **15**: 21-44.
Nitabach, M. N. and P. H. Taghert (2008). "Organization of the Drosophila circadian control circuit." Curr Biol **18**(2): R84-93.
Panda, S., M. P. Antoch, B. H. Miller, A. I. Su, A. B. Schook, M. Straume, P. G. Schultz, S. A. Kay, J. S. Takahashi and J. B. Hogenesch (2002). "Coordinated Transcription of Key Pathways in the Mouse by the Circadian Clock." Cell **109**(3): 307-320.
Parisky, K. M., J. Agosto, S. R. Pulver, Y. Shang, E. Kuklin, J. J. Hodge, K. Kang, X. Liu, P. A. Garrity, M. Rosbash and L. C. Griffith (2008). "PDF cells are a GABA-responsive wake-promoting component of the Drosophila sleep circuit." Neuron **60**(4): 672-682.
Perkins, L. A., H. S. Shim and N. Perrimon (2009). Initial TRiP stock collection.
Ranjbaran, Z., L. Keefer, E. Stepanski, A. Farhadi and A. Keshavarzian (2007). "The relevance of sleep abnormalities to chronic inflammatory conditions." Inflammation Research **56**(2): 51-57.
Rey, G., F. Cesbron, J. Rougemont, H. Reinke, M. Brunner and F. Naef (2011). "Genome-wide and phase-specific DNA-binding rhythms of BMAL1 control circadian output functions in mouse liver." PLoS Biol **9**(2): e1000595.
Scheiermann, C., Y. Kunisaki and P. S. Frenette (2013). "Circadian control of the immune system." Nat Rev Immunol **13**(3): 190-198.
Schmid, B., C. Helfrich-Forster and T. Yoshii (2011). "A new ImageJ plug-in "ActogramJ" for chronobiological analyses." J Biol Rhythms **26**(5): 464-467.
Schroder, E. A., M. Lefta, X. Zhang, D. C. Bartos, H. Z. Feng, Y. Zhao, A. Patwardhan, J. P. Jin, K. A. Esser and B. P. Delisle (2013). "The cardiomyocyte molecular clock, regulation of Scn5a, and arrhythmia susceptibility." Am J Physiol Cell Physiol **304**(10): C954-965.
Silver, A. C., A. Arjona, M. E. Hughes, M. N. Nitabach and E. Fikrig (2012a). "Circadian expression of clock genes in mouse macrophages, dendritic cells, and B cells." Brain Behav Immun **26**(3): 407-413.

Silver, A. C., A. Arjona, W. E. Walker and E. Fikrig (2012b). "The circadian clock controls toll-like receptor 9-mediated innate and adaptive immunity." Immunity **36**(2): 251-261.

Stone, E. F., B. O. Fulton, J. S. Ayres, L. N. Pham, J. Ziauddin and M. M. Shirasu-Hiza (2012). "The circadian clock protein timeless regulates phagocytosis of bacteria in Drosophila." PLoS Pathog **8**(1): e1002445.

Storch, K. F., O. Lipan, I. Leykin, N. Viswanathan, F. C. Davis, W. H. Wong and C. J. Weitz (2002). "Extensive and divergent circadian gene expression in liver and heart." Nature **417**(6884): 78-83.

Watson, F. L., R. Puttmann-Holgado, F. Thomas, D. L. Lamar, M. Hughes, M. Kondo, V. I. Rebel and D. Schmucker (2005). "Extensive diversity of Ig-superfamily proteins in the immune system of insects." Science **309**(5742): 1874-1878.

Wulff, K., S. Gatti, J. G. Wettstein and R. G. Foster (2010). "Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease." Nat Rev Neurosci **11**(8): 589-599.

Xu, K., Justin R. DiAngelo, Michael E. Hughes, John B. Hogenesch and A. Sehgal (2011). "The Circadian Clock Interacts with Metabolic Physiology to Influence Reproductive Fitness." Cell Metabolism **13**(6): 639-654.

Zhang, R., N. F. Lahens, H. I. Ballance, M. E. Hughes and J. B. Hogenesch (2014). "A circadian gene expression atlas in mammals: Implications for biology and medicine." Proc Natl Acad Sci U S A.

Chapter 4: *Achilles* regulates the rhythmicity of immune system in *Drosophila*

Abstract

Circadian rhythms are the daily oscillations of almost all aspects of life, including behavior, metabolism, and physiology. At the molecular level, it is regulated by core clock genes as well as clock-controlled genes (CCGs). Core clock genes compose the TTFL (transcriptional translational feedback loop) that maintains the 24-hr rhythm. CCGs, the molecular outputs of the circadian clock, orchestrate cellular, metabolic, and behavioral rhythms directly or indirectly. *Achilles* (*Achl*), a novel clock-controlled gene, has been shown to play an important role in the regulation of the immune system in *Drosophila melanogaster*. However, it is not clear if *Achl* also regulates the rhythmicity of immune response. Here we examined the role of *Achl* in the regulation of immune circadian rhythms by knocking down *Achl,* and we checked the immune response upon bacterial infection. We found that in *Achl* RNAi flies, the rhythmicity in both survival and sensitivity of immune response gene expression upon infection are disrupted, suggesting that *Achl* does regulate the immune response. In addition, we profiled the CCGs that lose rhythmicity upon *Achl* knock-down. These candidates genes will help understand the signaling cascade that *Achl* makes use of to send the regulatory signals from the brain to the immunological tissues.

Introduction

The earth rotates around its axis once in roughly 24 hours, giving us days and nights. Most organisms have developed an endogenous timing system to predict and adapt to this environmental change. Named circadian rhythms, organisms display all kinds of rhythmic physiological processes such as locomotor activity. For example, the immune system is circadianly regulated from insects to mammals (Lee and Edery 2008, Labrecque and Cermakian

2015), as both mice and flies show rhythmic survival upon infection. These rhythmic physiological processes are regulated both at molecular level and organismal level.

At molecular level, the circadian clock is composed of core clock genes as well as clock controlled genes (CCGs). The core clock genes compose and reinforce a self-sustaining transcriptional/translational feedback loop (TTFL) that finishes one cycle about twenty four hours. This core clock architecture is conserved between flies and mice. Taking the core clock in flies for example, two transcription factors, CLOCK (CLK) and CYCLE (CYC) drive the expression of *period* (per) and *timless (tim)*, which in turn inhibit the activity of CLOCK and CYCLE upon translation. Once inhibited, the level of PER and TIM will decrease, eventually releasing the blockage and starting a new cycle of *per* and *tim* expression. This self-sustaining TTFL cycles about every twenty-four hours and forms the basis of circadian clocks. Meanwhile, CLK and CYC, as transcription factors, drive the rhythmic expression of thousands of other genes, some of these genes are regulatory proteins that could drive the rhythmic expression of further downstream genes. All these circadian genes, directly or indirectly regulated by the core clock, are named clock controlled genes (CCGs). CCGs are not part of the core clock timing system, instead, these are genes that are related to specific rhythmic physiological processes (Panda et al. 2002, Hardin 2005, Allada and Chung 2010).

At an organismal level, the principal oscillator in *Drosophila* is composed of about 150 neurons, including dorsal lateral neurons (LNd), small and large ventral lateral neurons (lLNv and sLNv), and dorsal neurons (DN1, DN2 and DN3) (Nitabach and Taghert 2008, Allada and Chung 2010). These neurons receive light signals from (1), the visual system, including compound eyes, ocelli, and Hofbauer-Buchner eyelets and (2) CRYPTOCHROME (CRY), a blue-light photoreceptor that is expressed in many clock neurons. Upon receiving timing signals, clock neurons will

synchronize the timing cue and transmit the synchronized timing signals to other brain regions and peripheral tissues. Output signals are transmitted through neuronal connections and secretary factors. The principal oscillator thus ultimately coordinates rhythmic behavior, metabolism and physiology in peripheral tissues through this signal transmitting system. However, it is not completely clear how this signaling transmitting system works. Here using *Drosophila* immune system as a study model, we found that *AchI*, a CCG that is expressed in the brain, is a key factor that transmits circadian signals to the fat body, the major immunological tissue in flies to orchestrate the rhythmicity of the immune system in *Drosophila.* We identified CCGs downstream of *AchI* using high-throughput RNA-sequencing, these CCGs are candidate genes mediating signaling transmission from the brain to the fat body.

Profiling candidate downstream CCGs regulated by *AchI*

There is a conserved RNA binding domain within ACHL protein, suggesting that *AchI* is a regulatory protein that may regulate the rhythmic expression of further downstream CCGs. we thus investigated if *AchI* regulates downstream CCGs and profiled CCGs that lost their rhythmicity after *AchI* RNAi.

Starting at CT24, both control and *AchI* RNAi flies were collected every two hours for consecutive two days on both sexes. Fly heads were chosen for sample collection because there is both brain and fat body in the head. The core clock resides in the brain, and the immune system resides in the fat body. Sequencing libraries were generated from prepared RNA samples with in line control DNAs and quality control was performed with conventional methods. After sequencing with Illumina Hiseq platform, the raw sequencing results were mapped to the *Drosophila* genome and transcriptome with RUM algorithm (Grant et al. 2011), the rhythmicity of

the expressed transcripts were analyzed using JTK_CYCLE algorithm (Hughes et al. 2010). It turned out that on average 15 million reads per sample were obtained from RNA-sequencing (Table 4.1). Among these reads, 86.7% were uniquely mapped to the genome/transcriptome.

We first looked at the expression of the core clock genes. In agreement with previous observations, the core clock genes maintain their rhythms (Figure 4.1). We also looked at the expression pattern of *Achl*. Though *Achl* also maintains its rhythmicity upon RNAi, its amplitude damped significantly. The peak (highest) expression level of *Achl* expression in *Achl* RNAi flies is about the same level as the trough (lowest) expression in control flies, suggesting that this maintained yet dramatically damped rhythmicity in *Achl* is not strong enough to maintain its regulation towards downstream CCGs (Figure 4.2), thus this RNA-seq data can be used to detect CCGs that are downstream of *Achl*.

We hypothesized that *Achl* regulated-CCGs are genes that are rhythmic in control flies but lose the rhythmicity in *Achl* RNAi flies. To identify these *Achl* regulated-CCGs, we ran JTK_CYCLE and compared the rhythmicity pattern of all transcripts that are expressed in control and *Achl* RNAi flies (Table 4.2). In males, there are similar numbers of CCGs detected in both control and *Achl* RNAi flies. In females, there are dramatically fewer CCGs detected in *Achl* RNAi flies, suggesting that there might be a sex specific difference in *Achl* regulation towards immune system. Because we use males for all our functional assays, we decided to focus on the males in terms of CCGs detection.

As shown in Figure 4.3, we picked p value of < 0.0001 as the rhythmic threshold, and p value of > 0.001 as the non-rhythmic threshold. Choosing different thresholds here is to avoid some of the false positives that are still rhythmic in RNAi flies but the statistics is just around the rhythmic threshold. Consequently, 124 transcripts were found to maintain their rhythmicity and 92

Table 4.1. RNA-seq alignment statistics

Sequencing ID	Total reads (millions)	% aligned	unique reads (millions)	% uniquely aligned
Female *Achl* RNAi CT24	15.44	95.80%	13.26	82.33%
Female *Achl* RNAi CT26	20.95	96.10%	19.06	87.46%
Female *Achl* RNAi CT28	18.88	94.40%	17.05	85.27%
Female *Achl* RNAi CT30	15.4	96.80%	13.48	84.81%
Female *Achl* RNAi CT32	9.34	91.40%	7.7	75.34%
Female *Achl* RNAi CT34	7.54	87.90%	7.16	83.47%
Female *Achl* RNAi CT36	21.66	94%	18.99	82.45%
Female *Achl* RNAi CT38	19.23	95%	18.4	90.97%
Female *Achl* RNAi CT40	9.99	93.50%	8.67	81.21%
Female *Achl* RNAi CT42	17.03	93.40%	15.03	82.44%
Female *Achl* RNAi CT44	13.3	97.60%	11.77	86.44%
Female *Achl* RNAi CT46	13.54	97.90%	12.64	91.43%
Female *Achl* RNAi CT48	11.79	97.30%	11.37	93.87%
Female *Achl* RNAi CT50	13.6	97.10%	13.02	93%
Female *Achl* RNAi CT52	12.39	96.10%	11.86	92.06%
Female *Achl* RNAi CT54	14.38	88.60%	13.43	82.82%
Female *Achl* RNAi CT56	24.89	98.10%	23.66	93.28%
Female *Achl* RNAi CT58	18.37	98.10%	17.66	94.33%
Female *Achl* RNAi CT60	19.08	88.60%	17.62	81.91%
Female *Achl* RNAi CT62	17.39	86.20%	15.74	78.10%

Female *AchI* RNAi CT64	14.55	98%	13.8	93.02%
Female *AchI* RNAi CT66	16.94	97.60%	16.38	94.47%
Female *AchI* RNAi CT68	11.47	98.10%	11.1	94.99%
Female *AchI* RNAi CT70	20.91	98.20%	20.08	94.30%
Female *Control* CT24	18.61	97.30%	16.75	87.59%
Female *Control* CT26	20.2	96.50%	18.45	88.20%
Female *Control* CT28	13.82	92.80%	12.24	82.18%
Female *Control* CT30	15.8	97.20%	14.08	86.72%
Female *Control* CT32	14.6	97.70%	13.85	92.73%
Female *Control* CT34	15.8	98%	14.31	88.76%
Female *Control* CT36	14.32	95.30%	12.41	82.70%
Female *Control* CT38	20.81	88.90%	18.21	77.82%
Female *Control* CT40	16.06	95.70%	15.04	89.70%
Female *Control* CT42	16.63	95.90%	14.82	85.46%
Female *Control* CT44	20.41	98.10%	18.26	87.87%
Female *Control* CT46	11.26	96.40%	10.22	87.56%
Female *Control* CT48	13.98	97.80%	13.49	94.49%
Female *Control* CT50	18.32	98.10%	17.79	95.37%
Female *Control* CT52	17.73	94.50%	16.96	90.43%
Female *Control* CT54	14.2	97.90%	13.7	94.51%
Female *Control* CT56	14.27	97.90%	13.83	94.90%
Female *Control* CT58	17.8	97.90%	17.12	94.22%
Female *Control* CT60	15.43	96.80%	14.79	92.82%
Female *Control* CT62	13.26	94.70%	12.64	90.30%

Female *Control* CT64	13.8	97.30%	13.16	92.75%
Female *Control* CT66	16.71	98.10%	15.97	93.79%
Female *Control* CT68	14.64	97.80%	14.1	94.23%
Female *Control* CT70	16.85	97.90%	16.05	93.26%
Male *Achl* RNAi CT24	12.19	96.20%	10.64	83.97%
Male *Achl* RNAi CT26	20.06	95.80%	17.46	83.42%
Male *Achl* RNAi CT28	22.98	96.90%	19.96	84.16%
Male *Achl* RNAi CT30	19.15	96.50%	16.02	80.78%
Male *Achl* RNAi CT32	11.43	96.50%	9.58	80.94%
Male *Achl* RNAi CT34	15.89	96%	13.03	78.81%
Male *Achl* RNAi CT36	15.77	96.70%	13.55	83.12%
Male *Achl* RNAi CT38	12.45	97.10%	10.55	82.28%
Male *Achl* RNAi CT40	14	96%	12.4	85.12%
Male *Achl* RNAi CT42	15.42	96.50%	13.57	84.99%
Male *Achl* RNAi CT44	16.47	96%	14.6	85.11%
Male *Achl* RNAi CT46	16.58	96.40%	14.41	83.77%
Male *Achl* RNAi CT48	12.96	96.50%	11.29	84.18%
Male *Achl* RNAi CT50	14.83	96.70%	12.7	82.79%
Male *Achl* RNAi CT52	14.94	96.50%	12.63	81.63%
Male *Achl* RNAi CT54	15.39	96.20%	13.31	83.23%
Male *Achl* RNAi CT56	12.95	96.30%	11.3	84.03%
Male *Achl* RNAi CT58	14.26	96.10%	12.87	86.75%
Male *Achl* RNAi CT60	20.02	96.70%	17.84	86.18%
Male *Achl* RNAi CT62	14.22	97.30%	12.47	85.30%

Male *Achl* RNAi CT64	14.38	96.60%	12.98	87.17%
Male *Achl* RNAi CT66	17.05	94.70%	14.94	83.03%
Male *Achl* RNAi CT68	19.87	96.80%	17.51	85.35%
Male *Achl* RNAi CT70	15.48	96.70%	13.7	85.60%
Male *Control* CT24	15.76	96.60%	14.18	86.92%
Male *Control* CT26	20.3	97.10%	17.86	85.51%
Male *Control* CT28	13.42	95.80%	12.13	86.62%
Male *Control* CT30	13.06	96.20%	11.62	85.55%
Male *Control* CT32	13.17	97.10%	11.37	83.85%
Male *Control* CT34	18.6	96.50%	16.53	85.81%
Male *Control* CT36	18.14	97.50%	15.62	84.02%
Male *Control* CT38	16.49	97.50%	13.76	81.44%
Male *Control* CT40	14.98	95.20%	13.81	87.80%
Male *Control* CT42	18.4	94.50%	17.09	87.86%
Male *Control* CT44	18.79	96.70%	17.12	88.12%
Male *Control* CT46	15.25	94.10%	14.14	87.33%
Male *Control* CT48	16.49	96.70%	14.71	86.30%
Male *Control* CT50	18.64	95.90%	16.82	86.54%
Male *Control* CT52	24.1	95.90%	21.67	86.27%
Male *Control* CT54	16.2	96.40%	14.17	84.36%
Male *Control* CT56	12.17	97.10%	10.91	87.07%
Male *Control* CT58	11.74	96.70%	10.67	87.92%
Male *Control* CT60	14.08	97.10%	12.5	86.34%
Male *Control* CT62	12.81	97.10%	11.22	85.10%

Male *Control* CT64	10.27	97.10%	9	87.29%
Male *Control* CT66	13.2	97.10%	11.55	85%
Male *Control* CT68	15.21	97%	13.42	85.63%
Male *Control* CT70	13.4	97.30%	11.41	82.96%

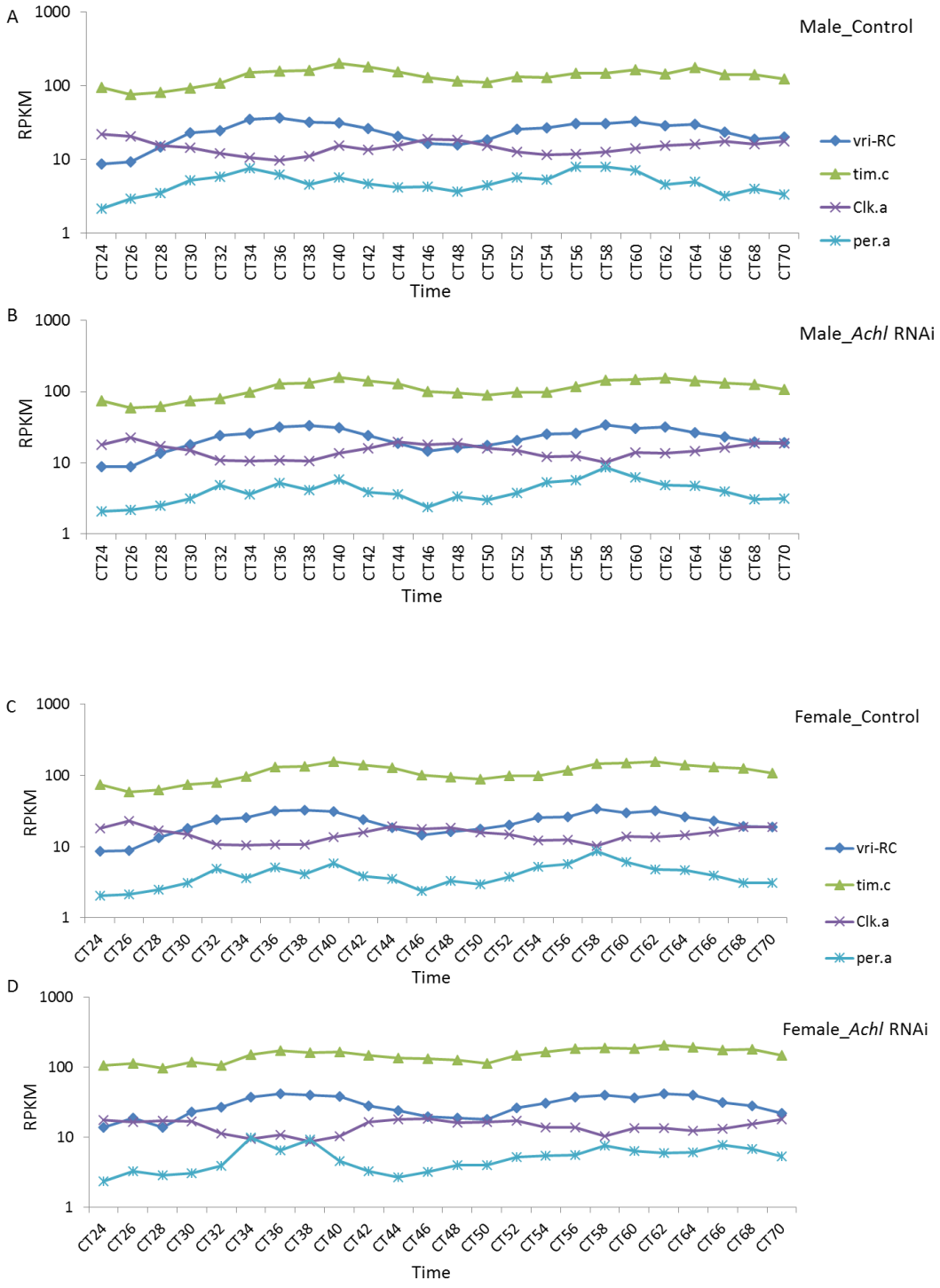
This table shows the RNA-seq statistics for all samples collected for RNA-seq, including two genotypes: *Achl* RNAi and Control. For each genotype we collected both sexes. And for each sex of that genotype, we collected 24 samples in a 2-hour resolution.

Table 4. 2. JTK_CYCLE analysis statistics

Number of transcripts	Female Control	Female RNAi	Male Control	Male RNAi
total	34830	34585	34917	34948
p <0.001 (q-value of 0.058 in male control sample)	379	112	550	569

These are the statistics of all RNA-seq data after running JTK-CYCLE algorithm.

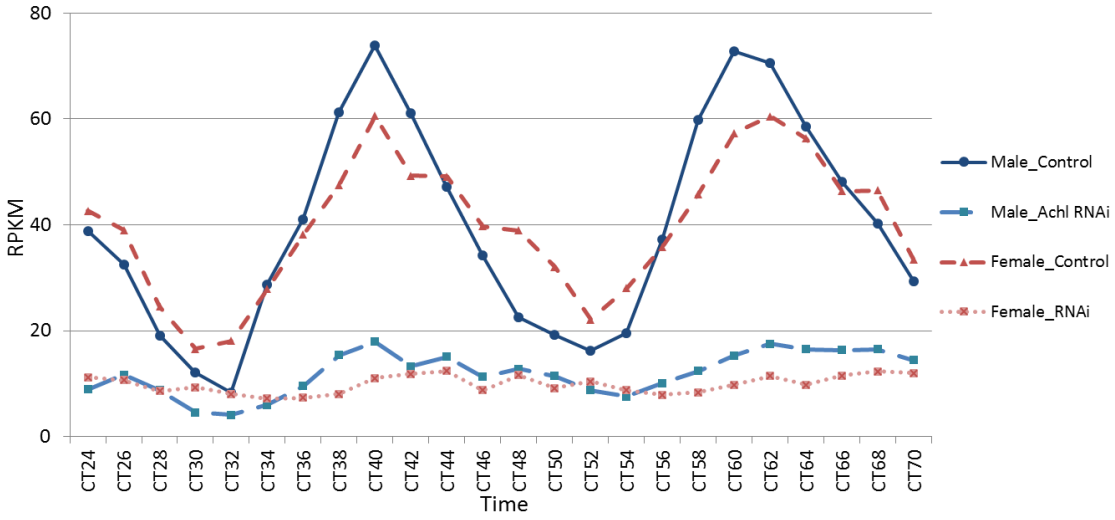
Figure 4.1. Core clock genes maintain their rhythmicity upon *Achl* RNAi



The expression pattern of four core clock genes, *vri, tim, Clk* and *per* in male (A & B) and female

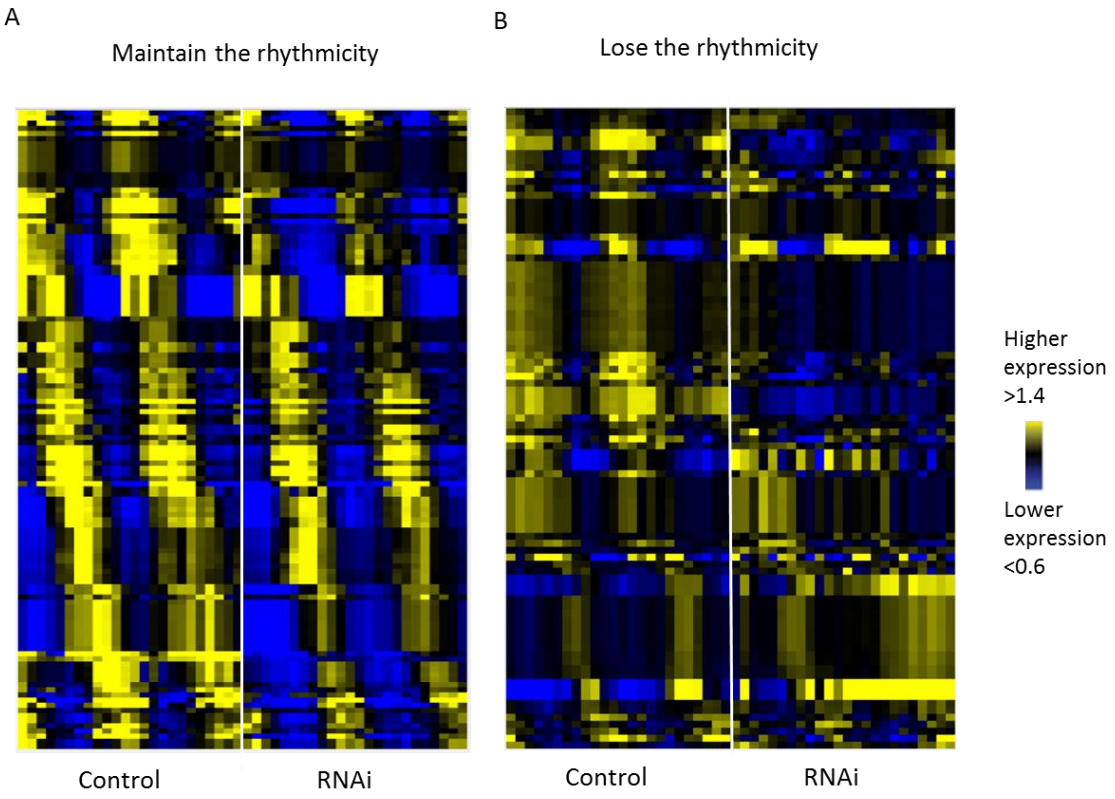
(C & D) flies from RNA-seq data confirming that *Achl* is not part of the core clock.

Figure 4.2. *Achl* expression is greatly knocked down by RNAi



The expression profile of *Achl* in both females and males of *Achl* RNAi and control flies from RNA-seq. *Achl* maintains its rhythmicity in RNAi flies. However, the overall expression level in RNAi flies is very low.

Figure 4.3. CCGs that either maintain or lose their rhythmicity upon Achl RNAi



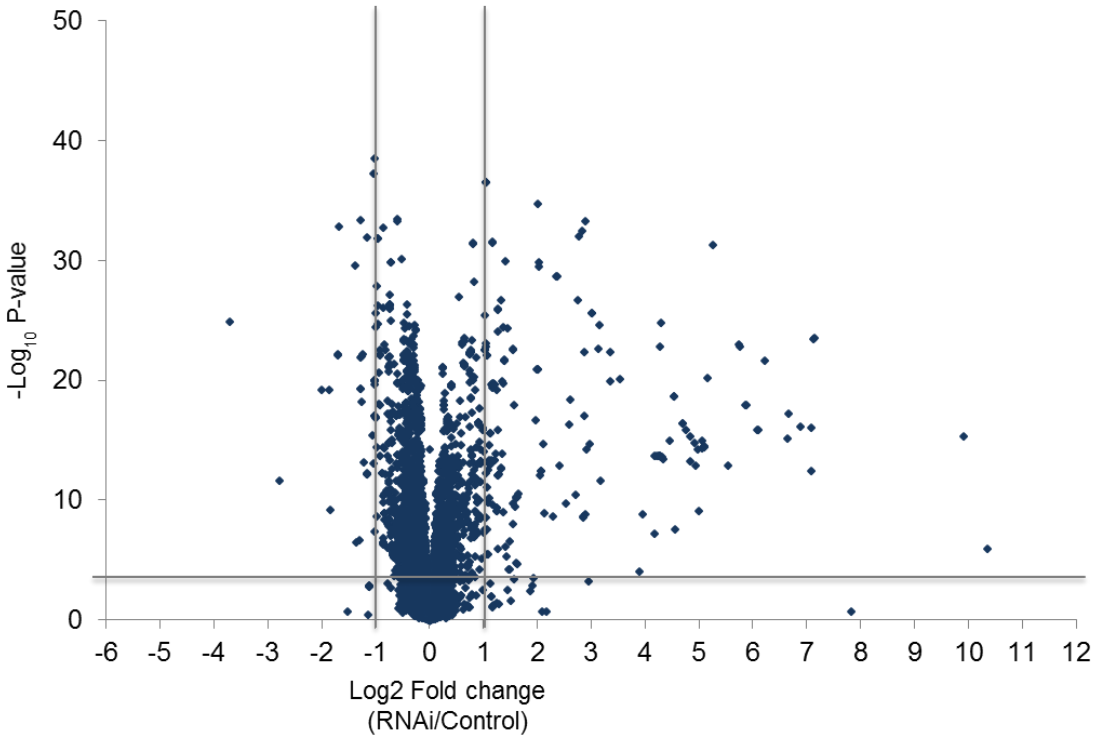
Heatmap of transcripts that maintain (A) or lose (B) their rhythmicity upon *Achl* knock-down. Each row indicates one transcript, and each volume indicates one time point. Yellow indicates higher expression, and blue indicates lower expression. The bottom lines indicate the genotype: blue line indicates GFP control, and orange line indicates *Achl* RNAi. The white line splits each heatmap according to the genotypes. The expression values are median-normalized.

transcripts were found to lose their rhythmicity upon *Achl* RNAi. The transcripts that maintain their rhythmicity are the core clock genes and CCGs that play other physiological roles unrelated to *Achl*. The transcripts that lose their rhythmicity are promising candidate genes that may be downstream of *Achl* and mediate the rhythmic immune response. For example, JHL-21 is an amino acid antiporter that have been shown to play a role in the immune response combating fungal infection (Jin et al. 2008). Nevertheless, we acknowledge that some of these candidates are likely to be false positives due to the false discovery in this kind of high throughput analysis. Further functional assay is required to determine if they play any *Achl* dependent role in the regulation of immune system.

Achl regulates immune response genes at all time points

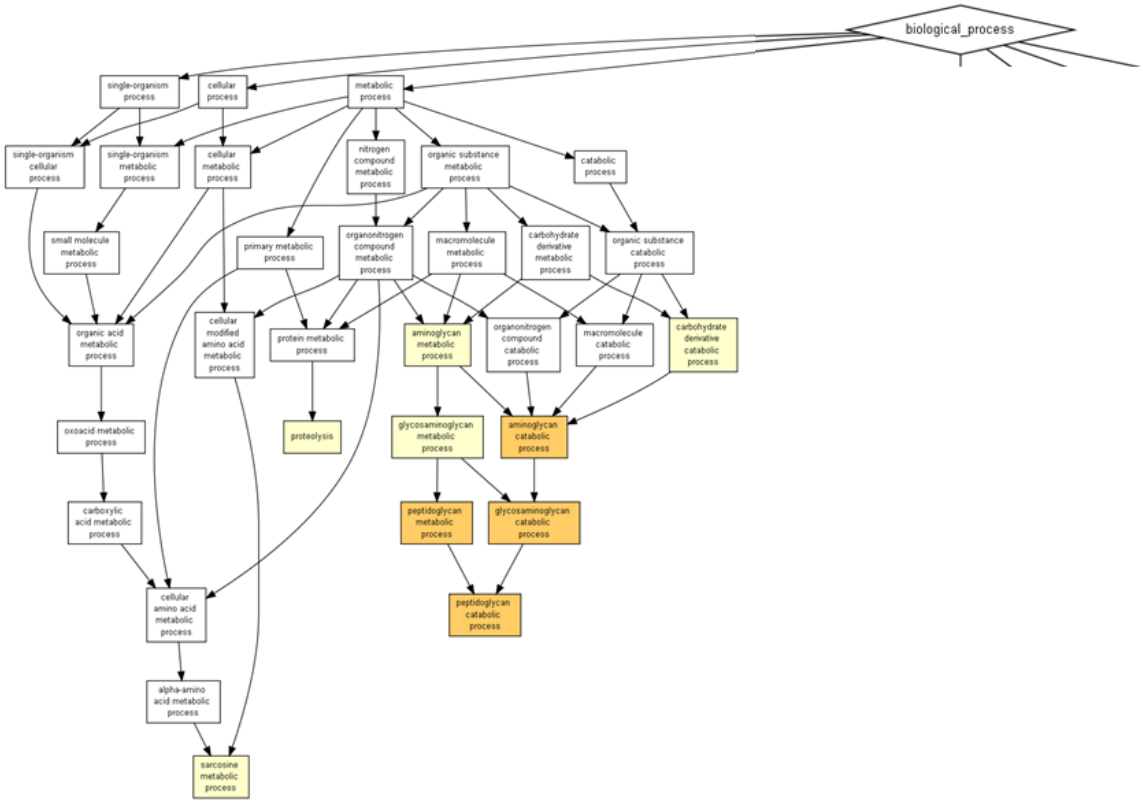
The RNA-seq data analyzed in Chapter 3 were obtained from the samples collected at ZT2, when the ACHL protein level is presumably highest. Consequently that RNA-seq data can only provide information limited to that time point. By analyzing RNA-seq data obtained here from samples collected at multiple times of the day, we will be able to find genes with altered expression at different time points. This finding may provide insights into the regulation of other physiological processes, like metabolism. Thus we analyzed all samples collected over the forty eight hours with two-way ANOVA analysis. As shown before, there are more genes being upregulated in *Achl* RNAi flies (Figure 4.4), confirming that *Achl* regulates the immune system in a repressive manner. As shown in Figure 4.5, the most strongly enriched GO terms are still immune related, validating the specificity of *Achl* in regulating the immune system. We also found a few general glycan metabolic/catabolic processes that are metabolism related, supporting our previous statement that there is an altered metabolic status as the trade-off for having a constitutively activated immune system (See detailed description of enriched GO terms

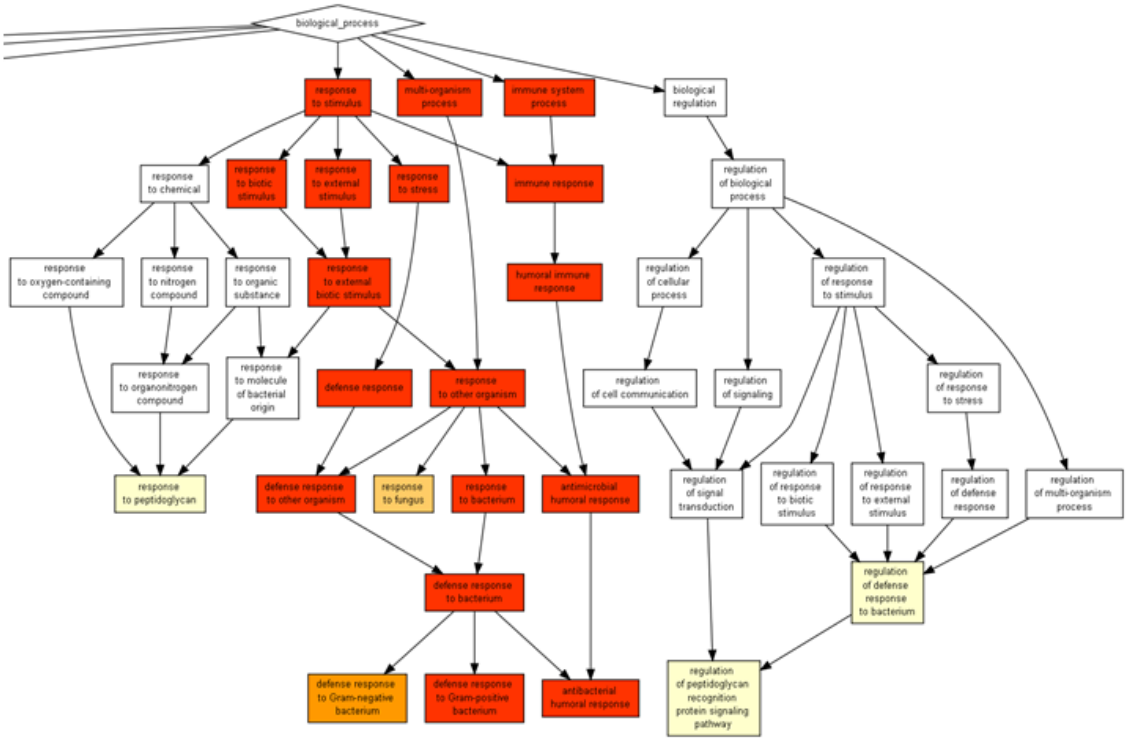
Figure 4.4. Volcano plot showing genes differentially expressed in *Achl* RNAi flies



Volcano plot of the RNA-seq data showing genes differentially expressed in *Achl* RNAi flies. Each dot represent a single transcript. Gray dotted lines indicate a p-value threshold of <0.001 (correspond to a q-value of and fold change threshold of 2X. For each genotype, data includes all 24 samples collected at multiple time points.

Figure 4.5. GO analysis reveals enrichment of immune related processes in *Achl* RNAi flies





GO analysis of differentially expressed genes in *AchI* RNAi flies. The figure is split into two part here to optimize visualization. There are two major groups of enriched processes. The left half is the group of metabolic processes and the right half is the group of immune response processes. Different color scales indicate different p-values of enrichment: White: $p > 10^{-3}$; light yellow: $10^{-3} < p < 10^{-5}$; light orange: $10^{-5} < p < 10^{-7}$; orange: $10^{-7} < p < 10^{-9}$; red: $p < 10^{-9}$. The connective arrow lines indicate that the GO terms of the arrow-head side are specialized from one general term of the other side. A detailed description of each enriched term is shown in Table 4.3.

Table 4.3. Detailed GO enrichment for Figure 4.10

GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)
GO:0050830	defense response to Gram-positive bacterium	1.96E-25	1.33E-21	37.97 (7827,35,106,18)
GO:0051707	response to other organism	1.80E-24	6.08E-21	9.79 (7827,249,106,33)
GO:0043207	response to external biotic stimulus	3.46E-24	7.80E-21	9.59 (7827,254,106,33)
GO:0009607	response to biotic stimulus	3.46E-24	5.85E-21	9.59 (7827,254,106,33)
GO:0009617	response to bacterium	5.53E-24	7.49E-21	12.76 (7827,162,106,28)
GO:0006952	defense response	4.82E-21	5.44E-18	8.48 (7827,270,106,31)
GO:0051704	multi-organism process	4.95E-21	4.79E-18	7.66 (7827,318,106,33)
GO:0019731	antibacterial humoral response	9.37E-19	7.93E-16	46.64 (7827,19,106,12)
GO:0098542	defense response to other organism	3.00E-18	2.25E-15	9.56 (7827,193,106,25)
GO:0042742	defense response to bacterium	1.10E-17	7.42E-15	11.28 (7827,144,106,22)
GO:0009605	response to external stimulus	8.08E-17	4.97E-14	5.01 (7827,531,106,36)
GO:0006955	immune response	4.22E-16	2.38E-13	9.56 (7827,170,106,22)
GO:0006959	humoral immune response	1.44E-15	7.51E-13	18.16 (7827,61,106,15)
GO:0002376	immune system process	1.07E-14	5.16E-12	7.68 (7827,221,106,23)
GO:0019730	antimicrobial humoral response	2.97E-13	1.34E-10	20.14 (7827,44,106,12)
GO:0050896	response to stimulus	1.83E-11	7.75E-09	2.65 (7827,1309,106,47)

Detailed GO enrichment table. This is a full Gene Ontology enrichment table generated by GOrilla software available online (Eden et al. 2007, Eden et al. 2009). The corresponding figure is shown in Figure 4.5.

As shown from the GOrilla website, Enrichment (N, B, n, b) is defined as follows:

N: the total number of genes; B: the total number of genes associated with a specific GO term; n: the number of genes in the top of the user's input list or in the target set when appropriate; b: the number of genes in the intersection.

and enrichment statistics in Table 4.3).

AchI regulates the rhythmicity of survival upon *S.aureus* infection

AchI regulates the immune response. The mRNA expression of *AchI* is robustly rhythmic. And the immune system is rhythmic. These three pieces of facts made it reasonable to hypothesize that *AchI* regulates the rhythmicity of the immune system. To test this hypothesis, we examined the ability of flies to combat infection in *AchI* RNAi and control flies at multiple times of the day by infecting flies with pathological bacteria *S.aureus* and monitoring the overall survival pattern after infection.

The needle mediated infection method that we used in Chapter 3 is convenient and fast, however, as shown in Figure 4.6A, it is hard to control the precise infecting bacterial load, thus making it difficult to do the multi-time-point comparisons. To minimize the variation that this inoculation method may introduce, we set up a nanoject III mediated infection assay. Nanoject III is the most advanced programmable nanoliter injector that allows accurate injection of small volume of liquid (bacterial solutions in our case). As shown in Figure 4.6B, nanoject III mediated infection is excellent in terms of reproducibility.

We started with *wildtype* Canton S (CS) flies. Starting at ZT4, flies of both genotype were infected with *S.aureus* at six different times of day. As shown in Figure 4.7, the median survival of infected flies show a rhythmic pattern that peaks around ZT20. This is in agreement with previous publication (Lee and Edery 2008), suggesting that our system is valid.

We then used this Nanoject III based infection system to test the median survival of *AchI* RNAi and control flies. As shown in Figure 4.8, the rhythm of *AchI* RNAi flies altered significantly,

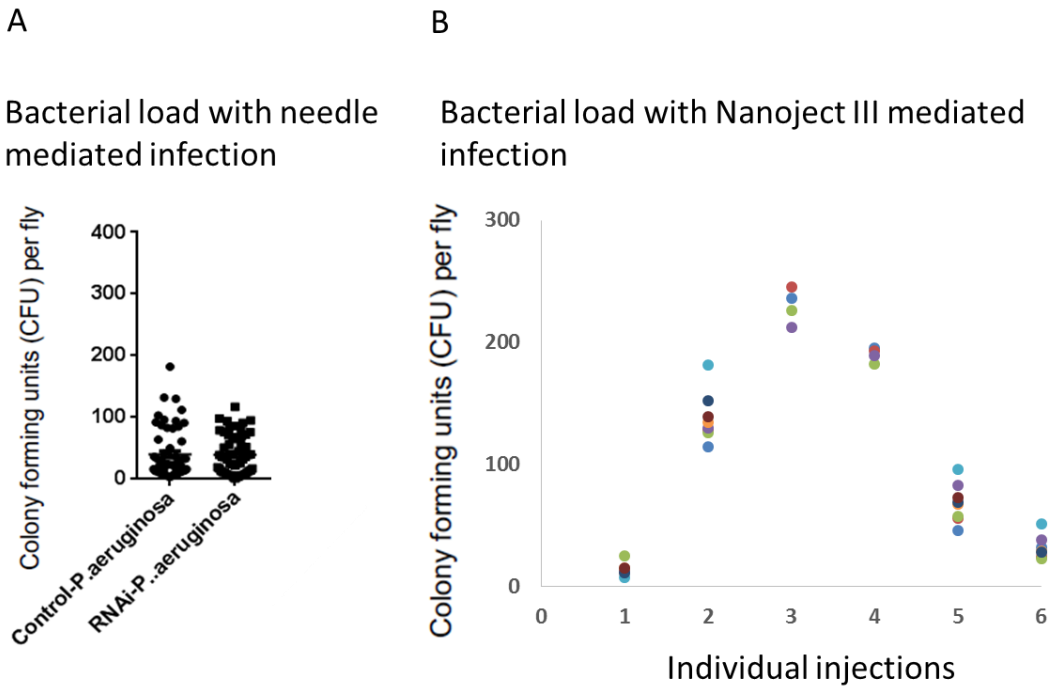
suggesting that *Achl* does regulate the rhythmicity of survival upon infection. Notably, the alteration pattern agrees with *Achl* expression pattern. The most significant changes happen in the first part of the day, where ACHL level is putatively high, suggesting that this alteration in survival is due to *Achl* RNAi.

Achl regulates the sensitivity of flies in inducing the expression of immune response genes

In addition to rhythmic survival upon infection, previous report showed that several immune response genes showed a rhythmic sensitivity upon infection (Lee and Edery 2008). Here we also examined this aspect in both control and *Achl* RNAi flies. Lipopolysaccharides (LPS), a major component of the outer membrane of the gram negative bacterial, can elicit strong immune response in flies. It was thus used to inoculate the flies in our assay due to its consistency and convenience.

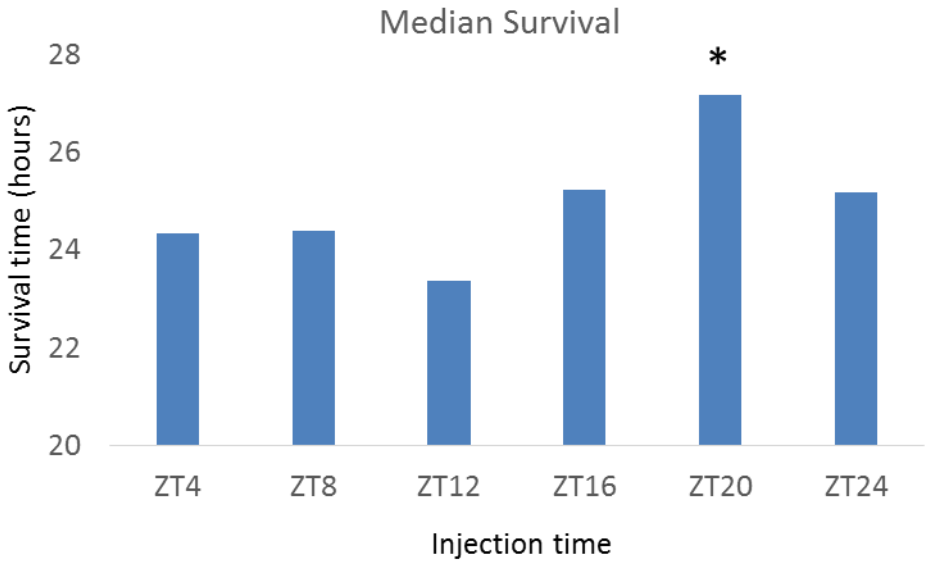
First, we examined the dose effect of LPS in inducing the immune response gene expression. Expression of immune response genes was examined with qPCR. As shown in Figure 4.9, LPS induces the expression of immune response genes in a dose dependent manner. We chose 5 ng as further infection dose because this is the dose with which *Mtk* expression reaches its plateau. Meanwhile, we found minimal death for these LPS injected flies, this is also why we only use LPS for immune response gene expression analysis.

Figure 4. 6. Nanoject III mediated infection has less variation in initial bacterial load



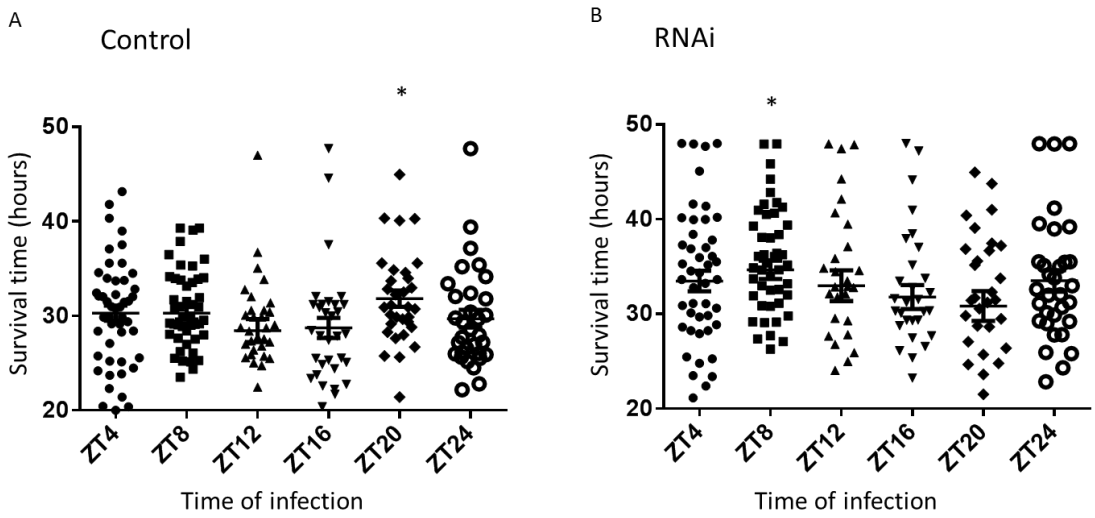
A comparison of two infection procedures. A is showing data from needle mediate infection, the data is adapted from Figure 3.10C. B is showing data from Nanoject III mediated infection, six experiments were performed with different bacterial load.

Figure 4.7. CS flies show rhythmic survival upon infection



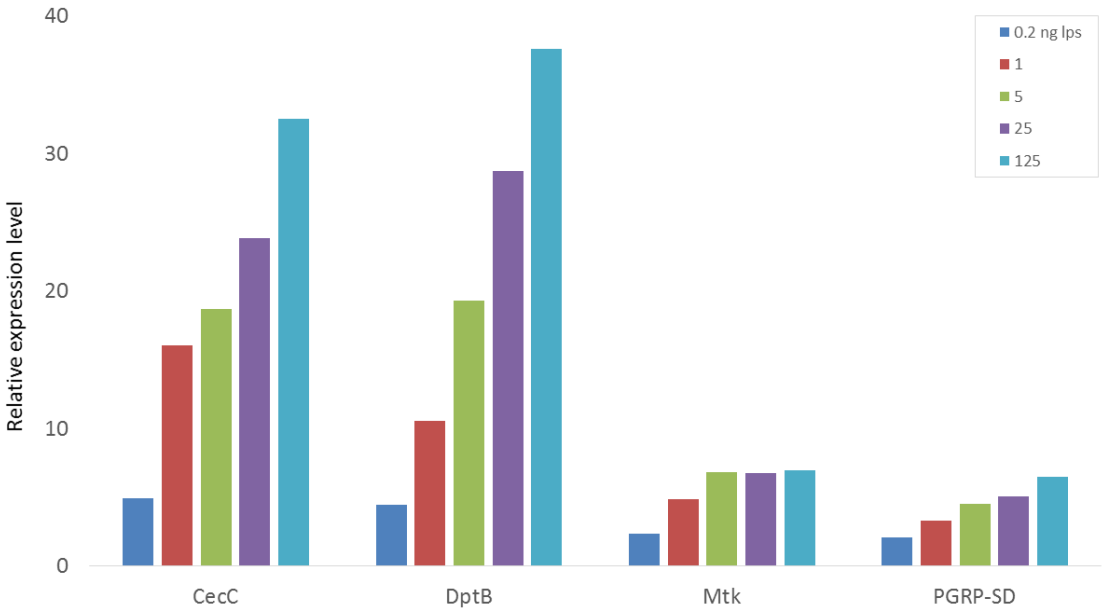
CS flies were infected at multiple times of the day to test the Nanoject III based system. CS flies show rhythmic survival upon infection after *S.aureus* infection at multiple times of the day. N = 15-16 for each genotype. *: p-value <0.05 when comparing the peak with surroundings using Student's T-test.

Figure 4.8. *Achl* RNAi flies have an altered survival rhythm upon infection



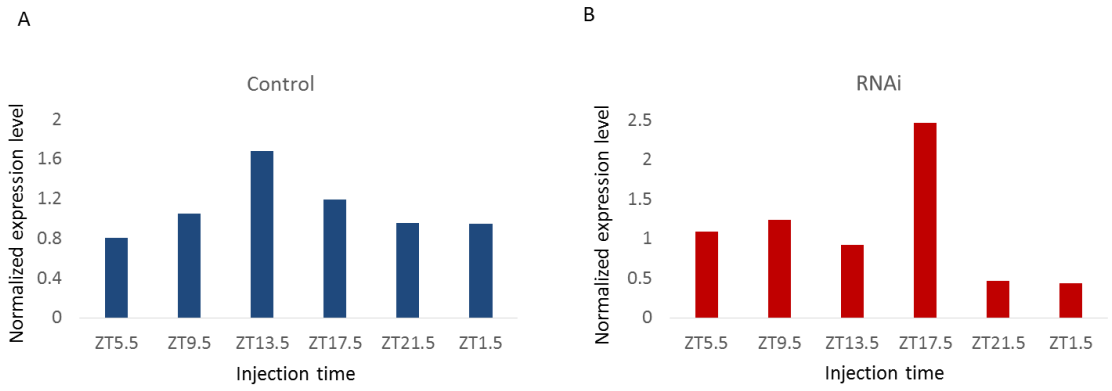
The median survival time of control (A) and *Achl* RNAi (B) flies after *S.aureus* infection at multiple times of the day suggesting that *Achl* RNAi flies have an altered survival rhythm upon infection. Each spot represents an individual fly. N = 15 - 16 for each replicate, and there are 2 - 3 replicates for each time point. *: $p<0.05$ when comparing the peak with surroundings using Student's T-test.

Figure 4.9. LPS induces the expression of immune response genes in a dose-dependent manner



LPS induce the expression of immune response genes in a dose-dependent manner. Four immune response genes, *CecC, DptB, Mtk* and *PGRP-SD* were examined. All samples were collected 6 hours post LPS infection. qPCR data were normalized with the expression value 0 hours post infection.

Figure 4.10. *Dro* lost its rhythmicity in the sensitivity of expression upon infection



Dro lost its rhythmicity in the sensitivity of expression upon infection. The expression level is detected six hours post infection by qPCR assay. (A). In control flies, the induced expression of *Dro* is rhythmic. JTK p-value: 0.1. The rhythmicity is comparable to the core clock gene *timeless* under the same resolution (Figure 3.1B). (B). In *AchI* RNAi flies, the induced expression of *Dro* lost its rhythmicity. JTK p-value: 1. qPCR data were normalized with the median expression value of each genotype.

We then investigated the sensitivity of both control and *AchI* RNAi flies in inducing the expression of immune response genes six hours after infection. As shown in Figure 4.10, *Dro*, a previously identified AMP that shows rhythmicity in its sensitivity upon infection, is rhythmic in control flies. However, this rhythmicity is gone in *AchI* RNAi flies, suggesting that *AchI* also regulates the sensitivity of flies in inducing the expression of immune response genes.

In which cell type would *AchI* play its role of regulating the immune system?

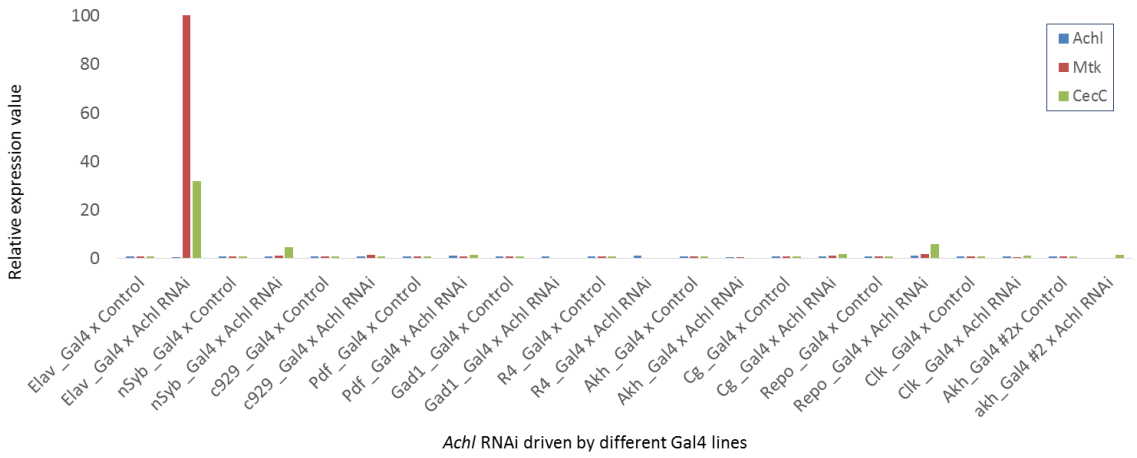
In all previous RNAi strategies, UAS (upstream activation sequence)-Gal4 system, a powerful genetic tool to manipulate tissue-specific expression of certain genes is used to drive *AchI* RNAi. Briefly, there is an *Elav-Gal4 Drosophila* line that drives the expression of Gal4 specifically in neurons. In addition, there is a *UAS-Dcr2* fragment inserted into the genome to enhance the efficiency of RNAi in this fly line. There is another *UAS-AchI RNAi Drosophila* line that has a UAS promoter followed by *AchI* RNAi sequence inserted into the genome. UAS promoter is only activated after GAL4 binding, thus this *UAS-AchI* RNAi element will only get activated when GAL4 protein is present. By crossing these two parental lines, neuron-specific *AchI* RNAi will be achieved in the F1 offspring.

We used *Elav-Gal4* to drive the knock-down of *AchI*, this Gal4 line drives *AchI* RNAi in the nervous system. However, it is not clear within which area, or which cell type would *AchI* play its role as an immune system regulator. To answer this question, we performed a UAS-Gal4 system based screening using multiple specialized Gal4 lines. When performing UAS-Gal4 system based screening, we realized that most Gal4 lines do not have a UAS-Dcr2 fragment inserted into their genome. To determine if missing *UAS-Dcr2* would affect the RNAi efficiency, we compared *AchI* RNAi efficiency using either *Elav-Gal4* alone or *Elav-Gal4* and *UAS-Dcr2*. It

turned out that both fly lines were effective in knocking down *AchI*. We thus decided to perform the screening without *UAS-Dcr2* fragment.

As shown in Figure 4.11, in our screening, Elav-gal4 turned out to be the only line that drives dramatic expression of immune response genes. Even another neuron-specific driver, *nSyb-gal4* did not drive such dramatic expression of immune response genes. There are three possibilities: (1) other driver lines could not drive *AchI* RNAi because they do not carry a *UAS-Dcer2* element as our *Elav-gal4* line does. Even though we have tested and confirmed that both Elav-gal4 alone and *Elav-gal4; UAS-Dcr2*, we cannot rule out the possibility that other gal4 lines need the help of *UAS-Dcr2* to process *AchI* RNAi. This possibility can be determined by adding *UAS-Dcr2* into the screening system using genetic tools. (2) We did not include the specific cell type driver in our screening, and even though both *nSyb-gal4* and *Elav-gal4* are recognized as neuron specific driver, there may be slight difference in their expression pattern, thus it is the cells that have *Elav* expression but not *nSyb* where *AchI* resides in and plays its role. This possibility will be determined by performing *in situ* hybridization or immunostaining using both *AchI* and *Elav/nSyb* probes/antibodies. (3) There is a slight increase of immune response gene expression in Repo-gal4 driving *AchI* RNAi flies. *Repo-gal4* is a glia specific marker, and *Elav-gal4* line was reported to be expressed in glia cells during early developmental stage (Berger et al. 2007). It is possible that it is actually that early stage glia cells that express *AchI* and have a long-lasting effect on the immune system. This possibility can also be determined by performing *in situ* hybridization or immunostaining of *AchI.* Meanwhile, we can analyze our RNA-seq data to determine if the expression level of glia specific genes is altered in *AchI* RNAi flies.

Figure 4.11. Screening to identify in which cells does *AchI* play its role in regulating the immune system



Cross different Gal4 lines with *AchI* RNAi and control to determine if driving *AchI* knock-down in that specific cell type is sufficient to regulate the immune system using qPCR assay.

Discussion

The circadian clock drives tissue specific expression of rhythmic mRNAs to regulate physiological processes, including the immune system. In chapter 3 we have revealed that *Achl*, a CCG, could regulate the expression of immune response genes, making us hypothesize that *Achl* also orchestrates the rhythmicity of immune system. To test this hypothesis, we infected both control and *Achl* RNAi flies at multiple times of the day and found that the rhythmicity in both survival and sensitivity of immune response gene expression upon infection is disrupted in *Achl* RNAi flies, confirming that *Achl* does play an essential role in the regulation of the rhythmicity of immune system (Figure 4.9 & 4.10).

There are limited numbers of immune response genes with a steady-state circadian expression in flies (McDonald and Rosbash 2001, Keegan et al. 2007, Xu et al. 2011). This is different from what happens in mice, where many immune response genes have been shown to be rhythmically expressed (Keller et al. 2009). Based on our findings, I would hypothesize that *Achl* actually plays a role in gating the expression of immune response genes upon infection at certain times of the day. This also agrees with our data that the sensitivity of immune response gene expression upon infection is disrupted in *Achl* RNAi flies.

Achl is not expressed in the fat body, where the majority of immune response genes are being expressed. Instead, it is expressed and plays its role in the brain. How does *Achl* regulate the gene expression in distant tissues? Here we identified CCGs that have disrupted rhythms upon knocking down *Achl* using high-throughput RNA-seq. These CCGs are potentially downstream of *Achl*. Whether any of them play a role in the signaling transmission of the immune system is not determined yet. Combining functional assay and ACHL RNA immunoprecipitation (RIP) may

help understand the details of this signaling cascade.

RNAi screening data have not yet confirmed which cells must express *Achl* to regulate immune function. As mentioned in the previous text, we will add UAS-Dcr2 element into our UAS-Gal4 screening system to make sure that the RNAi works. The slight increase of representative immune response genes in glia driven *Achl* RNAi (Repo-Gal4 line) flies is interesting. Glia are known to play important roles in regulating the immune system (Petersen et al. 2012, Cao et al. 2013). However, from our RNA-seq data, we did not find any difference in the expression of glia specific genes, nor genes known to play a role in the immune regulation in the glia, suggesting that *Achl* may regulate the immune response in a glia independent pathway. Even though earlier research suggested that Elav-Gal4 may be expressed early in glia cells, whether or not this *Achl* mediated immune regulation is through the glia is not clear yet. We are on the way of setting up Elav-gal4; repo-gal80 line driving *Achl* RNAi. The existence of repo-gal80 will prevent *Achl* knock-down in the glia, thus give us a clue if knocking down *Achl* in early glia cells in the early developmental glia cells contribute to the elevated immune response. In addition, in collaboration with Dr. Chen at McGill University, we are performing *Achl in situ* hybridization in the brain to determine its expression pattern to see in exactly which cells *Achl* are being expressed. We can then using corresponding Gal4 lines to drive *Achl* RNAi there and detect the immune system.

Materials and Methods

Bacterial stocks and culture

A frozen glycerol stock was freshly streaked onto a LB plate and grown overnight at 37 °C. A single colony was picked from this plate and grown in 2 ml LB media overnight. Afterwards a

subculture was made in 20 ml LB. The culture was harvested and added with 20% glycerol after overnight growth. Bacterial mixture were then aliquoted into 1ml aliquots in 1.5 ml Eppendorf tubes for storage at -80 degree. Before infection, frozen bacterial were taken out, centrifuged and suspended with fresh LB. After two and half hours of incubating in 37 degree shaking incubator, bacteria were collected, centrifuged, washed with PBS and serially diluted into a final OD of 0.04 for experimental infection.

Infection and survival assay

Survival rate assay protocol was performed as described previously (Apidianakis and Rahme, 2009). Briefly, male flies 1 to 2 days after eclosion were collected, raised in the incubator on standard food (Genesee Scientific, San Diego, CA) at 25 °C in LD conditions until 5 to 7 days after eclosion. At multiple zeitgeber time points, prepared bacterial solution will be filled with glass capillaries which will were then inserted into the Nanoject III head. 40nl of bacterial solution will then be injected into the thorax of individual flies. Each injured flies were moved into DAM (*Drosophila* activity monitor) system vials containing fresh food. The survival will be automatically monitored at 25 °C under LD condition. For each time point and genotype, n = 15 - 16.

The glass capillaries

We used micropipette puller P87 to generate the glass capillaries, with the following parameters: " Heat = 780, Pull = 0, Vel = 20, Time = 0". We break the tips of capillaries before use to generate sharp tiny openings.

Fly stocks

Flies were maintained on standard food (Genesee Scientific, San Diego, California) at 25 °C in 12 hour light: 12 hour dark (LD) conditions. Humidity was maintained at roughly 50%. *c929-Gal4* flies were obtained from Paul Taghert Lab at Washington University in St. Louis. All other fly stocks used were acquired from the Bloomington stock center:

All other experiments were performed as stated in Chapter 3

Bibliography

Allada, R. and B. Y. Chung (2010). "Circadian organization of behavior and physiology in Drosophila." Annu Rev Physiol **72**: 605-624.

Berger, C., S. Renner, K. Luer and G. M. Technau (2007). "The commonly used marker ELAV is transiently expressed in neuroblasts and glial cells in the Drosophila embryonic CNS." Dev Dyn **236**(12): 3562-3568.

Cao, Y., S. Chtarbanova, A. J. Petersen and B. Ganetzky (2013). "Dnr1 mutations cause neurodegeneration in Drosophila by activating the innate immune response in the brain." Proceedings of the National Academy of Sciences of the United States of America **110**(19): E1752-E1760.

Eden, E., D. Lipson, S. Yogev and Z. Yakhini (2007). "Discovering motifs in ranked lists of DNA sequences." PLoS Comput Biol **3**(3): e39.

Eden, E., R. Navon, I. Steinfeld, D. Lipson and Z. Yakhini (2009). "GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists." BMC Bioinformatics **10**(1): 1-7.

Grant, G. R., M. H. Farkas, A. D. Pizarro, N. F. Lahens, J. Schug, B. P. Brunk, C. J. Stoeckert, J. B. Hogenesch and E. A. Pierce (2011). "Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM)." Bioinformatics **27**(18): 2518-2528.

Hardin, P. E. (2005). "The Circadian Timekeeping System of Drosophila." Current Biology **15**(17): R714-R722.

Hughes, M. E., J. B. Hogenesch and K. Kornacker (2010). "JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets." J Biol Rhythms **25**(5): 372-380.

Jin, L. H., J. Shim, J. S. Yoon, B. Kim, J. Kim, J. Kim-Ha and Y. J. Kim (2008). "Identification and functional analysis of antifungal immune response genes in Drosophila." PLoS Pathogens **4**(10): e1000168.

Keegan, K. P., S. Pradhan, J. P. Wang and R. Allada (2007). "Meta-analysis of Drosophila circadian microarray studies identifies a novel set of rhythmically expressed genes." PLoS Comput Biol **3**(11): e208.

Keller, M., J. Mazuch, U. Abraham, G. D. Eom, E. D. Herzog, H. D. Volk, A. Kramer and B. Maier (2009). "A circadian clock in macrophages controls inflammatory immune responses." Proc Natl

Acad Sci U S A **106**(50): 21407-21412.

Labrecque, N. and N. Cermakian (2015). "Circadian Clocks in the Immune System." J Biol Rhythms **30**(4): 277-290.

Lee, J.-E. and I. Edery (2008). "Circadian Regulation in the Ability of Drosophila to Combat Pathogenic Infections." Current Biology **18**(3): 195-199.

McDonald, M. J. and M. Rosbash (2001). "Microarray analysis and organization of circadian gene expression in Drosophila." Cell **107**(5): 567-578.

Nitabach, M. N. and P. H. Taghert (2008). "Organization of the Drosophila circadian control circuit." Curr Biol **18**(2): R84-93.

Panda, S., J. B. Hogenesch and S. A. Kay (2002). "Circadian rhythms from flies to human." Nature **417**(6886): 329-335.

Petersen, A. J., S. A. Rimkus and D. A. Wassarman (2012). "ATM kinase inhibition in glial cells activates the innate immune response and causes neurodegeneration in Drosophila." Proceedings of the National Academy of Sciences **109**(11): E656–E664.

Xu, K., Justin R. DiAngelo, Michael E. Hughes, John B. Hogenesch and A. Sehgal (2011). "The Circadian Clock Interacts with Metabolic Physiology to Influence Reproductive Fitness." Cell Metabolism **13**(6): 639-654.

Chapter 5: Conclusion and future directions

AchI is a clock-controlled gene

Our qPCR data, as well as previously published high-throughput microarray and RNA-seq data, show that *AchI* is rhythmically expressed at the mRNA level in *wildtype* flies but not clock disrupted flies (Figure 3.1 & 3.2). *AchI* has tandem E-box regions within its promoter region, ChIP-seq data also suggest that CLOCK and CYCLE binds directly to *AchI* promoter region (Abruzzi et al. 2011, Meireles-Filho et al. 2014), suggesting that *AchI* is downstream of the core clock. *AchI* knock-down has no effect on the rhythms of the core clock (Figure 3.6). And the locomotor activity is not affected in *AchI* RNAi flies (Figure 3.7), confirming that *AchI* is not part of the core clock, instead, it is a clock controlled gene that is downstream of the core clock genes.

AchI regulates the expression of immune response genes

To study the functional role of *AchI*, we generated *AchI* RNAi flies using UAS-GAL4 system that allows tissue specific interference of the gene of interest. Because *AchI* is highly expressed in the brain, we chose Elav-Gal4 that drives putative neuron specific *AchI* RNAi and got over 60% RNAi efficiency (Figure 3.4). We then examined the genes that are affected by knocking-down *AchI*. RNA-seq from fly head samples revealed that knock-down of *AchI* specifically results in activated expression of immune response genes (Figure 3.8), suggesting that *AchI* regulates the expression of immune response genes in a repressive manner. Furthermore, we found activated expression of immune response genes in not only brains, but also the body, suggesting that this is a systematic activation (Figure 3.9).

AchI regulates the ability to combat infection largely by increasing the resistance towards infection

Since knocking-down *AchI* causes upregulation of immune response genes, we hypothesized that *AchI* RNAi flies will also have a better survival after infection. To test this, we infected flies with pathological bacteria and examined the survival afterwards. In agreement with our hypothesis, *AchI* RNAi flies survive better than control flies (Figure 3.10A & 3.10B). Meanwhile, we found no significant difference in the level of immune response genes at 24 hours after infection in control and *AchI* RNAi flies (Figure 3.12D and Figure 3.12), suggesting that there is no increased ability to induce higher immune response gene expression in *AchI* RNAi flies, it is the increased steady-state immune response gene expression that contributes to the increased survival. To further determine if this better survival is due to increased resistance, or increased tolerance, we performed the colony formation unit assay. Flies immediately or 24 hours after infection were homogenized and plated to determine their bacterial load. While *AchI* RNAi and control flies carry the same initial bacterial load (Figure 3.11), there is a significant decrease in the bacterial load in *AchI* RNAi flies after 24 hours (Figure 3.10C), suggesting that this better survival is largely due to increased resistance.

AchI regulates the rhythmicity of survival upon infection

After confirming that *AchI* knock-down cause dramatic activation in the expression of immune response genes as well as increased survival at time point ZT2, we investigated if AchI also regulates the rhythmicity of the immune system. To test this, we set up the Nanoject III mediated infection assay to inject a precise initial bacterial load into fly body at multiple times of the day (Figure 4.6). It turned out that *AchI* RNAi flies do have an altered survival rhythm upon infection, and the alteration mainly focused on the early light-on times, when ACHL level is presumably highest (Figure 4.8). This agrees with the previous statement that *AchI* regulates the immune system in a repressive way.

AchI regulates the sensitivity of immune response gene expression upon infection

In addition to the rhythmic survival regulation, we also investigated the rhythmicity of the sensitivity of immune response gene expression upon infection. To do so, we injected LPS to induce the expression of immune response genes at multiple times of the day. Six hours afterwards, fly heads were collected and qPCR was performed to determine the expression of immune response genes. As shown in Figure 4.10, *Drosocin* (*Dro*) show a nice rhythm in its sensitivity towards infection at different times of the day. However, there is not such a rhythm in *AchI* RNAi flies, suggesting that *AchI* also regulates the rhythm of the sensitivity towards infection in flies. Notably, we did not find such a rhythm in the induced expression pattern of several other anti-microbial peptides, it is possible that it takes less than 6 hours for these genes to be induced, future experiments should be performed earlier after infection to test if that is the case.

Decreased lifespan and starvation resistance as trade-offs

While increasing resistance towards bacterial infection, flies with knocked-down *AchI* show a decreased overall lifespan under natural uninfected conditions (Figure 3.13). This physiological trade-off between immunity and lifespan, as shown in other organisms as well (Mills et al. 2010, Todesco et al. 2010, Schwenke et al. 2016), suggests the critical role circadian rhythms plays in making full use of resources to balance multiple demands in a time-dependent fashion. As demonstrated here, *AchI* represses the steady-state expression of immune response genes at certain times of the day to take best advantage of resources available and disrupting this balance is detrimental to flies.

AchI regulates downstream CCGs that may mediate the signaling cascade from the brain to the fat body

AchI is expressed in the brain while immune response genes are expressed in the immunological tissues, mainly the fat body. Therefore, there must be a signaling cascade that transmits the circadian signals from one tissue to the other. After confirming *AchI*'s role in regulating the immune system, especially the rhythmicity of the immune system, we investigated the underlying signaling cascade by collecting samples every two hours for two consecutive days, performing RNA-seq, and analyzing the rhythmicity of the expressed transcriptome. We identified 92 transcripts that code for CCGs. Losing their rhythmicity upon *AchI* RNAi, these CCGs are promising candidates that are downstream of *AchI* and mediate signaling from the brain to the fat body (Figure 4.3). Nevertheless, further functional assay is necessary to validate their roles.

It is not completely clear where does *AchI* regulate the immune system yet

In this project we mostly used Elav-gal4, a putative pan-neuronal driver to express *AchI* RNAi. We tried screening with more cell type specific drivers, however it did not work well (Figure 4.10). Possibilities and on-going optimization to the screening have been described in detail in Chapter 4. We have started to collaborate with Dr. Chen lab to do *in situ* hybridization to get more information about the cell type that expresses *AchI* and regulate the immune system.

Future directions

Here we described our project focusing on *AchI,* a novel gene that plays a fundamental role in regulating the immune system and its rhythmicity in *Drosophila*. In this project we focused on the

molecular side of the immune system, that is, the expression of immune response genes. We did not investigate the cellular side, like the phagocytosis process. Previous research has shown that the phagocytosis process in the *Drosophila* is also under the circadian regulation (Stone et al. 2012). It will be interesting to check if *AchI* also plays any role in the regulation of phagocytosis. Related to this, we did not test if the microbiota in flies is regulated by *AchI.*

Previous studies have shown the significance of post-transcriptional regulation in generating circadian rhythms while most studies still focus on the regulation at the transcriptional level. With its conserved putative RNA-binding domain, ACHL protein may serve as a nice model to study post-transcriptional regulation of circadian rhythms. To do so, tagged *AchI* transgenic flies and *AchI* mutant flies with RNA binding domain truncated can be made. The tagged fly line can be used to examine the expressing pattern of *AchI.* In addition, the tagged fly can be used to perform RNA-immunoprecipitation-seq (RIP-seq), which will tell us genes that are directly controlled by *AchI*. The truncation fly line will be used to examine the significance of this post-transcriptional RNA binding in regulating the rhythmicity of immune response.

We have found CCGs potentially regulated by *AchI*. However, how this is regulated and if this plays a role in the immune system is still not clear. Future functional work can be done to examine whether the candidate gene plays a role in this process. In combination with RIP-seq data, we may be able to decipher the signaling cascade from *AchI* to the rhythmic immune system. This finding will provide valuable insights in understanding the regulation of tissue specific peripheral physiological circadian rhythms.

Lastly, there are other orthologs that belong to the same sub-family with *AchI* in *Drosophila,* as well as homologs in mammals, it will be interesting to check if any of these proteins also show rhythmicity and play any role in rhythmic regulation of any physiological process. Actually there

are two homologs, *Larp6* and *Larp 7* that show a rhythmic mRNA expression in mice, though not in the immune system. It will be interesting to see if they also play any role in orchestrating any physiological process with a similar regulating mechanism.

Bibliography

Abruzzi, K. C., J. Rodriguez, J. S. Menet, J. Desrochers, A. Zadina, W. Luo, S. Tkachev and M. Rosbash (2011). "Drosophila CLOCK target gene characterization: implications for circadian tissue-specific gene expression." Genes Dev **25**(22): 2374-2386.

Meireles-Filho, A. C., A. F. Bardet, J. O. Yanez-Cuna, G. Stampfel and A. Stark (2014). "cis-regulatory requirements for tissue-specific programs of the circadian clock." Curr Biol **24**(1): 1-10.

Mills, S. C., A. Grapputo, I. Jokinen, E. Koskela, T. Mappes and T. Poikonen (2010). "Fitness trade-offs mediated by immunosuppression costs in a small mammal." Evolution **64**(1): 166-179.

Schwenke, R. A., B. P. Lazzaro and M. F. Wolfner (2016). "Reproduction–Immunity Trade-Offs in Insects." Annual review of entomology **61**: 239-256.

Stone, E. F., B. O. Fulton, J. S. Ayres, L. N. Pham, J. Ziauddin and M. M. Shirasu-Hiza (2012). "The circadian clock protein timeless regulates phagocytosis of bacteria in Drosophila." PLoS Pathog **8**(1): e1002445.

Todesco, M., S. Balasubramanian, T. T. Hu, M. B. Traw, M. Horton, P. Epple, C. Kuhns, S. Sureshkumar, C. Schwartz, C. Lanz, R. A. E. Laitinen, Y. Huang, J. Chory, V. Lipka, J. O. Borevitz, J. L. Dangl, J. Bergelson, M. Nordborg and D. Weigel (2010). "Natural allelic variation underlying a major fitness trade-off in Arabidopsis thaliana." Nature **465**(7298): 632-636.